

Statistical Calibration of qRT-PCR, Microarray and RNA-Seq Gene Expression Data with Measurement Error Models

Zhaonan Sun, Thomas Kuczek and Yu Zhu

*Address of the authors
Department of Statistics
Purdue University
250N University Street
West Lafayette, IN 47907-2066*

Abstract: The accurate quantification of gene expression levels is crucial for transcriptome study. Microarray has been used as a main platform for simultaneously interrogating thousands of genes in the past decade, and recently RNA-Seq has emerged as a promising alternative. The gene expression measurements obtained by microarray and RNA-Seq are however subject to various measurement errors. A third platform called qRT-PCR is acknowledged to provide more accurate quantification of gene expression levels than microarray and RNA-Seq, but it has limited throughput capacity. In this article, we propose to use a system of functional measurement error models to model gene expression measurements and calibrate the microarray and RNA-Seq platforms with qRT-PCR. Based on the system, a two-step approach was developed to estimate the biases and error variance components of the three platforms and calculate calibrated estimates of gene expression levels. The estimated biases and variance components shed light on the relative strengths and weaknesses of the three platforms and the calibrated estimates provide a more accurate and consistent quantification of gene expression levels. Theoretical and simulation studies were conducted to establish the properties of those estimates. The system was applied to analyze two gene expression data from the Microarray Quality Control (MAQC) and Sequencing Quality Control (SEQC) projects.

Keywords and phrases: Transcriptome Profiling, Gene Differential Expression, Comparative Calibration, Functional and Structural Parameters.

1. Introduction

The transcriptome of a cell is the entire set of RNA molecules or transcripts produced by DNA transcription under certain biological or environmental conditions. Systematic profiling of the transcriptome can not only provide a dynamic characterization of the cell's molecular constitution, but also shed light on gene functional annotation, regulatory mechanisms, and transcriptional networks underlying various biological processes of the cell. Since the mid-1990s, DNA microarray has served as the leading experimental platform for transcriptome study (Schena et al. [20]). Despite its huge success, microarray is known to

suffer from some limitations such as reliance on existing knowledge of transcript sequences, high level background noise and limited dynamic range of detection.

Recently, a new experimental platform called RNA-Seq, which is based on Next Generation Sequencing (NGS) technologies, has emerged as a promising alternative to microarray for transcriptome profiling. The massive parallel throughput capacity of RNA-Seq provides whole genome coverage as well as single nucleotide resolution. Therefore RNA-Seq overcomes major limitations of microarray (Wang, Gerstein and Snyder [24]). Data generated from RNA-Seq experiments still demonstrate excessive variability. A number of studies published in the literature found that RNA-Seq is subject to various types of bias and variation attributable to a variety of sources; see Schraga, Oren and Ast [21] for a comprehensive review. Therefore, similar to microarray data, RNA-Seq data need to be normalized for downstream transcriptome analysis.

A number of methods have been proposed in the literature for normalizing RNA-Seq data. The RPKM method proposed by Mortazavi *et al.* [16] is arguably the most widely used method, which summarizes RNA-Seq data into Reads Per Kilobase of exon model per Million mapped reads for measuring gene or transcript expression levels. The RPKM method is intuitive and easy to implement, however, it ignores biases and variations observed in RNA-Seq data. In order to improve upon the RPKM method, various statistical model-based methods have been proposed, including GPseq (Srivastava and Chen [23]), mseq (Li, Jiang and Wong [14]), and POME (Hu *et al.* [12]). Despite reported improvements of the newly proposed methods over RPKM, the problem of how to statistically characterize and further normalize RNA-Seq data has not been settled with satisfaction and demands further investigation (Mak [15]).

For gene expression quantification, a third platform called the quantitative real-time Reverse Transcription Polymerase Chain Reaction (qRT-PCR) can also be used. Acknowledged as the most reliable gene expression measurement technology, qRT-PCR is sensitive to low levels of gene expression, has a wide dynamic range of detection, and is highly reproducible (Bustin [6]). Nonetheless, qRT-PCR is not perfect and also subject to variations caused by various biological, technical, and experimental factors involved in qRT-PCR experiments (Bustin and Nolan [7]). The major disadvantage of qRT-PCR is that it is not a massive parallel technology, and thus the number of genes or transcripts that can be measured in a single qRT-PCR experiment is limited. This disadvantage prevents qRT-PCR from being used for large scale transcriptome studies. Currently, qRT-PCR is either used for detecting and quantifying a small number of specific transcript targets or as a gold standard for validating hits or findings from other high-throughput platforms such as microarray (Applied Biosystems [2]). These two uses of qRT-PCR have not fully realized the potential of the technology for genome-wide gene expression profiling.

The strengths and weaknesses of qRT-PCR, RNA-Seq and microarray can be summarized as follows. In terms of throughput capacity, RNA-Seq is the highest, microarray is the second, and qRT-PCR is the lowest; whereas, in terms of accuracy, the order from the highest to the lowest is qRT-PCR, RNA-Seq, and microarray. An ideal platform for transcriptome profiling should combine

the accuracy of qRT-PCR with the high-throughput capacity of RNA-Seq. Unfortunately, such a platform is not currently available yet. A natural question is whether it is possible to use statistical methods to combine the strengths of the three platforms and generate gene expression measurements of a higher quality at the genome-wide scale. The answer turns out to be positive, and the methodology that can be applied is *statistical calibration*.

Statistical calibration is typically used for the scenario where p instruments are available for measuring the same quantity. Among the instruments, one is more accurate but at the same time more expensive than the others. In order to make a less accurate instrument generate more reliable measurement results for general use, it needs to be calibrated by the most accurate instrument through statistical analysis. There are two different types of calibration, which are *absolute calibration* and *comparative calibration*. The former refers to the type of calibration performed when the most accurate instrument gives the true value of the targeted quantity, whereas the latter refers to the type of calibration performed when the most accurate instrument is also subject to measurement error. The literature on statistical calibration is primarily focused on absolute calibration, and a variety of statistical methods have been developed; see Osborne [17] for a comprehensive review. On the other hand, comparative calibration is more challenging than absolute calibration and is mainly discussed in the literature on measurement error models; see Fuller [10] and Cheng and Van Ness [8] for more thorough discussions.

The platforms for measuring gene expression levels using qRT-PCR, microarray, and RNA-Seq represent three different instruments, with the qRT-PCR platform being the most accurate and the most expensive. In this article, we propose to use a system of measurement error (ME) models to model gene expression measurements obtained by the qRT-PCR, microarray, and RNA-Seq platforms, and to further use the system to calibrate RNA-Seq and microarray measurements with qRT-PCR measurements. A two-step approach is used to estimate the parameters of the system of ME models, and the statistical properties of the resulting estimates are discussed. Through both theoretical and simulation studies, we show that the calibrated gene expression measurements are more consistent and accurate than those by any of the individual platforms. Furthermore, we apply the system of ME models to calibrate gene expression data generated by qRT-PCR, RNA-Seq, and microarray from both the Microarray Quality Control (MAQC) and Sequencing Quality Control (SEQC) projects (detailed information about the data is given in Section 4.1), and show that the resulting calibrated measurements provide more accurate quantification of gene expression levels and have more power in gene differential expression analysis.

The rest of the article is organized as follows. In Section 2, we define the system of measurement error (ME) models and describe the two-step approach for estimating the parameters in the system. In Section 3, we present some simulation results. In Section 4, we describe the gene expression data from the MAQC and SEQC projects and discuss the results from applying the system of ME models to analyze the data. Section 5 concludes the article with further discussion and possible future research. An Appendix is enclosed at the end of

this article.

2. Measurement Error Models and Calibration

Let \mathcal{T} denote the collection of all the genes or transcripts in the transcriptome under study. As discussed in the Introduction, RNA-Seq has the capacity to measure the expression levels of all genes in \mathcal{T} , microarray can target thousands of genes, and qRT-PCR will generally measure hundreds of genes at most in a single experiment. Let \mathcal{A} , \mathcal{B} , and \mathcal{C} be the collections of genes measured by qRT-PCR, microarray, and RNA-Seq, respectively. We assume that $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C} \subset \mathcal{T}$, and the cardinalities of \mathcal{A} , \mathcal{B} , and \mathcal{C} are n , m , and l , respectively. By the assumption, n genes (i.e. those in \mathcal{A}) are measured by all three platforms, $m - n$ genes (i.e. those in $\mathcal{B} - \mathcal{A}$) are measured by both microarray and RNA-Seq, and $l - m$ genes (i.e. those in $\mathcal{C} - \mathcal{B}$) are measured only by RNA-Seq. We further assume that commonly used platform-specific methods are used to process and normalize raw data generated by each platform to produce gene expression measurement data. Some examples are the PLIER method (Affymetrix Inc. [1]) for microarray, the RPKM method for RNA-Seq, and the delta-delta C_t method for qRT-PCR. Following the convention in transcriptome studies, the log-2 transformation is further applied to the normalized gene expression measurements of the three platforms, and the resulting values are referred to as the qRT-PCR, microarray and RNA-Seq gene expression measurements, respectively, in the rest of this article.

Each type of gene expression measurement data can be generated by one lab or multiple labs. Based on the number of labs involved in generating gene expression measurement data, we distinguish two scenarios, which are the single-lab scenario and the multi-lab scenario. In the single-lab scenario, each type of gene expression measurement data is generated by a single lab. Note that the labs in the single-lab scenario may be the same lab. In the multi-lab scenario at least one type of gene expression data is generated by more than one lab. The rationale for separating these two scenarios is that variability attributable to labs can be assessed under the multi-lab scenario but not under the single-lab scenario.

2.1. Single-lab Scenario

We assume each lab generates one expression measurement for each gene. When technical replicates are available within one lab, their average value is used as the expression measurement of a gene. Suppose that the genes in \mathcal{A} , $\mathcal{B} - \mathcal{A}$, and $\mathcal{C} - \mathcal{B}$ are labeled by integers from 1 to n , $n + 1$ to m , and $m + 1$ to l , respectively. For $1 \leq j \leq n$, let X_j denote the expression measurement of gene j by qRT-PCR. Similarly, let Y_j denote the expression measurement of gene j by microarray for $1 \leq j \leq m$, and Z_j the expression measurement of gene j by RNA-Seq for $1 \leq j \leq l$. Note that X_j , Y_j and Z_j are all in log-2 scales.

As discussed in the Introduction, gene expression measurements produced by the three platforms are all subject to measurement errors. We propose to use the following measurement error models to characterize X_j , Y_j and Z_j , respectively,

$$X_j = \mu_j + \epsilon_{1j} \quad (1 \leq j \leq n), \quad (2.1a)$$

$$Y_j = \alpha_2 + \beta_2 \mu_j + \epsilon_{2j} \quad (1 \leq j \leq m), \quad (2.1b)$$

$$Z_j = \alpha_3 + \beta_3 \mu_j + \epsilon_{3j} \quad (1 \leq j \leq l). \quad (2.1c)$$

Note that the above models are not simple linear models. All terms in the right-hand side of the equations are not observed, and there are cross terms of unknown quantities. Next, we will discuss the terms in these three models in detail and propose the estimation method.

In model (2.1a), μ_j is the true expression level of gene j , ϵ_{1j} is the measurement error due to the qRT-PCR platform. Depending on what normalization method is used, qRT-PCR can lead to either absolute quantitation or relative quantitation of a gene's expression level (Pfaffl [18, Chapter 3]). Therefore, the interpretation of μ_j depends on whether absolute or relative quantitation is used in experiment. For simplicity, in this article, we do not further distinguish these two quantitation methods and simply refer to μ_j as the true expression value of gene j . We assume that for $1 \leq j \leq n$, ϵ_{1j} 's are i.i.d. $N(0, \sigma_1^2)$. The variance of X_j is σ_1^2 , representing the reproducibility of the qRT-PCR platform. According to the model, the qRT-PCR measurement X_j is unbiased with respect to μ_j .

In model (2.1b), ϵ_{2j} is the measurement error due to the microarray platform. For $1 \leq j \leq m$, we assume ϵ_{2j} 's are i.i.d. $N(0, \sigma_2^2)$. The other two terms form the mean part of the model, i.e., $E(Y_{jr}) = \alpha_2 + \beta_2 \mu_j$. In other words, the mean of the microarray measurement of gene j is assumed to be a linear function of the true expression level μ_j , where α_2 and β_2 are the intercept and slope, respectively, with β_2 representing the scale of the microarray measurements relative to the qRT-PCR measurements. The variance of Y_j is σ_2^2 . When comparing the reproducibilities of the qRT-PCR and microarray platforms, Y_j needs to be transformed to $\tilde{Y}_j = (Y_j - \alpha_2)/\beta_2$, which is of the same scale as X_j . The variance of \tilde{Y}_j , which is σ_2^2/β_2^2 , is referred to as the reproducibility of the microarray platform.

In model (2.1c), the mean of the RNA-Seq measurement is also assumed to be a linear function of μ_j , i.e., $E(Z_j) = \alpha_3 + \beta_3 \mu_j$, and ϵ_{3j} is the measurement error due to the RNA-Seq platform. We assume that for $1 \leq j \leq l$, ϵ_{3j} 's are i.i.d. $N(0, \sigma_3^2)$. The slope β_3 represents the scale of RNA-Seq measurements relative to the qRT-PCR measurements. The variance of Z_j is σ_3^2 . Similar to the microarray platform, we refer to σ_3^2/β_3^2 as the reproducibility of the RNA-Seq platform.

Different measurement platforms generally have different dynamic ranges of detection. This is also the case for the qRT-PCR, microarray and RNA-Seq platforms. The models proposed above may only hold for a range that fits all three platforms. Therefore, when applying the models in practice, a proper range of expression levels needs to be used, and genes with extremely low or high expression levels need to be excluded. Measurement errors are typically heteroscedastic, that is, the variance of a measurement error depends on the

magnitude of the targeted quantity. Notice that the errors in the models above are assumed to homoscedastic. The justification for assuming homoscedastic errors is two-fold. First, the gene expression measurements X_j , Y_j , and Z_j are the log-2 values of the original measurements produced by the three platforms, respectively, and the log-2 transformation is known to mitigate heteroscedasticity. Second, as previously discussed, the proposed models will be applied to genes with expression levels within a certain proper range, which further alleviates the concern of heteroscedasticity. Nonetheless, when applying the models, diagnostic analysis always needs to be performed.

When μ_j 's in models (2.1a)-(2.1c) are assumed to be unknown fixed quantities, the measurement error models are said to be *functional*; and when μ_j 's are assumed to be i.i.d. random variables, the models are said to be *structural*. In this article, μ_j 's are gene expression values and considered fixed. Therefore, the models defined above are functional. Following the convention in the literature, we call the μ_j 's *incidental parameters*, and the other parameters (α_i , β_i for $2 \leq i \leq 3$, and σ_i^2 for $1 \leq i \leq 3$) *structural parameters*.

In general, when p measurement platforms or instruments are used to measure the same quantity, it requires p measurement error models to characterize the measurement results. For ease of discussion, the p measurement error models are said to form a system of ME models of order p . When the measurement error models are structural, the system is said to be structural; and when the models are functional, the system is said to be functional. Hence, models (2.1a)-(2.1c) defined above form a functional system of order 3, and any two of them form a functional system of order 2.

Parameter Estimation The statistical literature on measurement error models is primarily focused on systems of order 2. Reiersøl [19] showed that structural systems of order 2 are not identifiable under the normality assumption, unless additional information such as the ratio between the variances of the two measurement instruments is available. For structural systems of order higher than 2 (e.g. $p = 3$), Barnett [3] showed they become identifiable with no further information needed, and the maximum likelihood estimates (MLEs) of the structural parameters can be obtained.

Parameter estimation for functional systems of ME models is more difficult than that for structural systems due to the presence of the incidental parameters, μ_j 's. First of all, under the normality assumption on the involved measurement errors, the likelihood function of a functional system becomes unbounded (Kendall and Stuart [13]). Solari [22] further showed that the likelihood function of a functional system of order 2 does not have a local maximum. We have found that this is also generally true for functional systems of order 3. Therefore, the maximum likelihood method fails to produce proper estimates for both structural and incidental parameters of a functional system of ME models.

The relationship between structural and functional systems was discussed in the early literature on measurement error models. Much attention has been given to the connection between the identifiability of structural systems and the existence of proper estimates of the structural parameters in functional systems.

One major result states that when a structural system is identifiable, then the MLEs for the structural system are still proper estimates for the structural parameters of the corresponding functional system (Gleser [11]).

Based on the discussion above, a two-step approach can be used to estimate the parameters of the functional system of order 3 defined above for the qRT-PCR, microarray and RNA-Seq measurements. In the first step, the genes with measurements from all three platforms (i.e. genes in \mathcal{A}) are used to obtain the MLEs of the structural parameters under the assumption that these μ_j 's are i.i.d. $N(\mu, \sigma^2)$. In the second step, the estimates obtained in the first step are used in place of their corresponding structural parameters in the functional ME system, and then the generalized least squares method is used to obtain estimates of the incidental parameters (or gene expression levels).

Structural Parameters We use the expression measurements of the genes in \mathcal{A} to estimate the structural parameters. Let \bar{X} , \bar{Y} , and \bar{Z} be the averages of X_j , Y_j , and Z_j over all genes in \mathcal{A} . The sample variances and covariances for $\{X_j\}_{1 \leq j \leq n}$, $\{Y_j\}_{1 \leq j \leq n}$, and $\{Z_j\}_{1 \leq j \leq n}$ are denoted as S_{xx} , S_{yy} , S_{zz} , S_{xy} , S_{xz} , and S_{yz} , respectively. Following the first step of the two-step approach discussed previously, the estimates of the structural parameters are calculated and given as follows.

$$\hat{\beta}_2 = \frac{S_{yz}}{S_{xz}}, \quad \hat{\beta}_3 = \frac{S_{yz}}{S_{xy}}, \quad \hat{\alpha}_2 = \bar{Y} - \hat{\beta}_2 \bar{X}, \quad \hat{\alpha}_3 = \bar{Z} - \hat{\beta}_2 \bar{X}, \quad (2.2a)$$

$$\hat{\sigma}_1^2 = \left(S_{xx} - \frac{S_{xy}S_{xz}}{S_{yz}} \right), \quad \hat{\sigma}_2^2 = \left(S_{yy} - \frac{S_{xy}S_{yz}}{S_{xz}} \right), \quad \hat{\sigma}_3^2 = \left(S_{zz} - \frac{S_{yz}S_{xz}}{S_{xy}} \right). \quad (2.2b)$$

Note that when deriving the above estimates, the non-negativity constraints on the variance estimates were not enforced due to two considerations. First, when analyzing real data, the constraints are usually satisfied automatically, and thus the benefit of using more complicated estimates is minimal; second, when a constraint is violated, it is usually an indication that the underlying variance is negligible and thus the negative estimate can be automatically converted to zero.

For convenience, let $\boldsymbol{\theta} = (\alpha_2, \alpha_3, \beta_2, \beta_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)^\top$ be the vector of the structural parameters, and $\hat{\boldsymbol{\theta}}$ the estimate of $\boldsymbol{\theta}$. The following proposition establishes the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ under the functional system of ME models as n goes to ∞ .

Proposition 1. *Assume the functional system of ME models is true and the following limits exist,*

$$\bar{\mu} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mu_j, \quad \text{and} \quad \Delta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\mu_j - \bar{\mu})^2 > 0. \quad (2.3)$$

Then, as n goes to ∞ , $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \Gamma_{\boldsymbol{\theta}})$, where \xrightarrow{D} means converge in distribution, and $\Gamma_{\boldsymbol{\theta}}$ is the variance-covariance matrix with its explicit expression given in Appendix A.2.

The proof of Proposition 1 is outlined in Appendix A.1. The asymptotic property of $\hat{\boldsymbol{\theta}}$ is stated as n goes to ∞ . In practice, this property will hold approximately when n is sufficiently large. As will be seen in the simulation study in Section 3, $n \geq 150$ is large enough for the approximation to be sufficiently accurate. The assumptions on the mean and variance of μ_j 's in Proposition 1 appear to be reasonable, because μ_j 's are the true expression levels of genes in a given cell line or tissue sample.

Incidental Parameters How to best estimate the incidental parameter μ_j depends on whether gene j is in \mathcal{A} , $\mathcal{B} - \mathcal{A}$, or $\mathcal{C} - \mathcal{B}$. We consider these three cases separately. For each case, we follow the second step of the two-step approach discussed previously. First, we replace the structural parameters by their estimates, and then we apply the generalized least squares method to obtain the estimate of μ_j .

For each $j \in \mathcal{A}$, given the structural parameters, the measurement error models involving μ_j can be rewritten as a linear model of μ_j (Appendix B). Replacing the structural parameters $\boldsymbol{\theta}$ by their estimates $\hat{\boldsymbol{\theta}}$, and applying the generalized least squares method, the estimate for μ_j is

$$\hat{\mu}_j^{xyz} = \frac{X_j/\hat{\sigma}_1^2 + \hat{\beta}_2(Y_j - \hat{\alpha}_2)/\hat{\sigma}_2^2 + \hat{\beta}_3(Z_j - \hat{\alpha}_3)/\hat{\sigma}_3^2}{1/\hat{\sigma}_1^2 + \hat{\beta}_2^2/\hat{\sigma}_2^2 + \hat{\beta}_3^2/\hat{\sigma}_3^2}, \quad (2.4)$$

We call $\hat{\mu}_j^{xyz}$ the calibrated estimate of μ_j for $j \in \mathcal{A}$, because it incorporates all of the information in the three types of measurements available for μ_j . The quality of $\hat{\mu}_j^{xyz}$ depends on how accurate the estimates of the structural parameters are, which further depends n , the number of genes in \mathcal{A} . The properties of $\hat{\mu}_j^{xyz}$ as n goes to ∞ are given in the following proposition.

Proposition 2. *Assume the functional system of ME models is true, and the limits in (2.3) exist. For each $j \in \mathcal{A}$, as n goes to ∞ , the calibrated estimate $\hat{\mu}_j^{xyz}$ of μ_j asymptotically follows the normal distribution $N(\mu_j, \gamma_{\mathcal{A}})$, and the expectation and variance of $\hat{\mu}_j^{xyz}$ admit the following expansions:*

$$E(\hat{\mu}_j^{xyz}) = \mu_j + O(n^{-1}), \quad (2.5a)$$

$$\text{Var}(\hat{\mu}_j^{xyz}) = \gamma_{\mathcal{A}} + n^{-1}\omega_{\mathcal{A}}(\boldsymbol{\theta}) + O(n^{-2}), \quad (2.5b)$$

where $\gamma_{\mathcal{A}} = (1/\sigma_1^2 + \beta_2^2/\sigma_2^2 + \beta_3^2/\sigma_3^2)^{-1}$, and the explicit expression of $\omega_{\mathcal{A}}(\boldsymbol{\theta})$ is given in Appendix C.

From Proposition 2, $\hat{\mu}_j^{xyz}$ is an asymptotically unbiased estimate of μ_j as n goes to ∞ . The variance of $\hat{\mu}_j^{xyz}$ does not converge to zero asymptotically, instead, it converges to $\gamma_{\mathcal{A}}$. When n is small or moderate, the second term in the expansion of $\text{Var}(\hat{\mu}_j^{xyz})$, i.e. $n^{-1}\omega_{\mathcal{A}}(\boldsymbol{\theta})$, may not be negligible. When n is sufficiently large, it becomes negligible, and the variance of $\hat{\mu}_j^{xyz}$ can be approximated by $\gamma_{\mathcal{A}}$. By replacing the structural parameters in $\gamma_{\mathcal{A}}$ with their corresponding estimates, we obtain the estimated variance and standard error of $\hat{\mu}_j^{xyz}$, which are $\hat{\gamma}_{\mathcal{A}}$ and $\sqrt{\hat{\gamma}_{\mathcal{A}}}$, respectively.

Note that the qRT-PCR measurement for μ_j is X_j , which is an unbiased estimate of μ_j and its reproducibility is σ_1^2 . Because $\gamma_{\mathcal{A}}$ is less than σ_1^2 , $\hat{\mu}_j^{xyz}$ is more accurate than X_j when n is sufficiently large. Therefore, by combining the measurements from all three platforms, we can obtain a more accurate estimate of the expression level of gene j .

Genes in $\mathcal{B} - \mathcal{A}$ are measured by microarray and RNA-Seq but not by qRT-PCR. Therefore, only Y_j and Z_j are available for $j \in \mathcal{B} - \mathcal{A}$. For each $j \in \mathcal{B} - \mathcal{A}$, given the structural parameters, the measurement error models involving μ_j can be rewritten as a linear model for μ_j (Appendix B). Replacing the structural parameters θ with their estimates and applying the generalized least squares method, the estimate of μ_j is

$$\hat{\mu}_j^{yz} = \frac{\hat{\beta}_2 (Y_j - \hat{\alpha}_2) / \hat{\sigma}_2^2 + \hat{\beta}_3 (Z_j - \hat{\alpha}_3) / \hat{\sigma}_3^2}{\hat{\beta}_2^2 / \hat{\sigma}_2^2 + \hat{\beta}_3^2 / \hat{\sigma}_3^2}. \quad (2.6)$$

We call $\hat{\mu}_j^{yz}$ the calibrated estimate of μ_j for $j \in \mathcal{B} - \mathcal{A}$. The properties of $\hat{\mu}_j^{yz}$ as n goes to ∞ are presented in the following proposition.

Proposition 3. *Assume the functional system of ME models holds, and the limits in (2.3) exist. For $j \in \mathcal{B} - \mathcal{A}$, as n goes to ∞ , the calibrated estimate $\hat{\mu}_j^{yz}$ of μ_j asymptotically follows the normal distribution $N(\mu_j, \gamma_{\mathcal{B}-\mathcal{A}})$, and the mean and variance of $\hat{\mu}_j^{yz}$ admit the following expansions:*

$$E(\hat{\mu}_j^{yz}) = \mu_j + O(n^{-1}), \quad (2.7a)$$

$$\text{Var}(\hat{\mu}_j^{yz}) = \gamma_{\mathcal{B}-\mathcal{A}} + n^{-1} \omega_{\mathcal{B}-\mathcal{A}}(\theta) + O(n^{-2}), \quad (2.7b)$$

where $\gamma_{\mathcal{B}-\mathcal{A}} = (\beta_2^2 / \sigma_2^2 + \beta_3^2 / \sigma_3^2)^{-1}$, and the explicit expression of $\omega_{\mathcal{B}-\mathcal{A}}(\theta)$ is given in Appendix C.

From Proposition 3, $\hat{\mu}_j^{yz}$ is an asymptotically unbiased estimate of μ_j , and the asymptotic variance of $\hat{\mu}_j^{yz}$ is $\gamma_{\mathcal{B}-\mathcal{A}}$. When n is small or moderate, the term $n^{-1} \omega_{\mathcal{B}-\mathcal{A}}(\theta)$ in the expansion of the variance of $\hat{\mu}_j^{yz}$ is not negligible. When n is sufficiently large, $n^{-1} \omega_{\mathcal{B}-\mathcal{A}}(\theta)$ becomes negligible, and thus the standard error of $\hat{\mu}_j^{yz}$ can be approximated by $\sqrt{\hat{\gamma}_{\mathcal{B}-\mathcal{A}}}$, where $\hat{\gamma}_{\mathcal{B}-\mathcal{A}}$ is the estimate of $\gamma_{\mathcal{B}-\mathcal{A}}$. It is clear that $\gamma_{\mathcal{B}-\mathcal{A}}$ is larger than $\gamma_{\mathcal{A}}$, which implies that the calibrated estimate of μ_j based on the measurements of RNA-Seq and microarray is less accurate than the calibrated estimate based on those by all three platforms. Further compare $\hat{\mu}_j^{yz}$ with the microarray measurement Y_j and the RNA-Seq measurement Z_j . Because $\gamma_{\mathcal{B}-\mathcal{A}}$ is less than either of σ_2^2 / β_2^2 and σ_3^2 / β_3^2 , which are the reproducibilities of Y_j and Z_j , respectively, $\hat{\mu}_j^{yz}$ is more accurate than Y_j and Z_j .

Genes in $\mathcal{C} - \mathcal{B}$ are only measured by RNA-Seq but not by qRT-PCR and microarray. Therefore, only Z_j are available for $j \in \mathcal{C} - \mathcal{B}$. For each $j \in \mathcal{C} - \mathcal{B}$, given the structural parameters, the measurement error models involving μ_j

can be written as a linear model for μ_j (Appendix B). Replacing the structural parameters with their estimates, the estimate of μ_j is

$$\hat{\mu}_j^z = (Z_j - \hat{\alpha}_3) / \hat{\beta}_3. \quad (2.8)$$

We call $\hat{\mu}_j^z$ the calibrated estimate of μ_j for $j \in \mathcal{C} - \mathcal{B}$. The properties of $\hat{\mu}_j^z$ as n goes to ∞ are presented in the following proposition.

Proposition 4. *Assume the functional system of ME models holds, and the limits in (2.3) exist. For $j \in \mathcal{C} - \mathcal{B}$, as n goes to ∞ , the calibrated estimates $\hat{\mu}_j^z$ of μ_j asymptotically follows the normal distribution $N(\mu_j, \gamma_{\mathcal{C}-\mathcal{B}})$, and the expectation and variance of $\hat{\mu}_j^z$ admit the following expansions*

$$E(\hat{\mu}_j^z) = \mu_j + O(n^{-1}), \quad (2.9a)$$

$$\text{Var}(\hat{\mu}_j^z) = \gamma_{\mathcal{C}-\mathcal{B}} + n^{-1}\omega_{\mathcal{C}-\mathcal{B}}(\boldsymbol{\theta}) + O(n^{-2}), \quad (2.9b)$$

where $\gamma_{\mathcal{C}-\mathcal{B}} = (\beta_3^2/\sigma_3^2)^{-1}$, and the explicit expression of $\omega_{\mathcal{C}-\mathcal{B}}(\boldsymbol{\theta})$ is given in Appendix C.

From Proposition 4, $\hat{\mu}_j^z$ is an asymptotically unbiased estimate of μ_j . When n is small or moderate, the term $n^{-1}\omega_{\mathcal{C}-\mathcal{B}}(\boldsymbol{\theta})$ in the expansion of the variance is not negligible. When n is sufficiently large, $n^{-1}\omega_{\mathcal{C}-\mathcal{B}}(\boldsymbol{\theta})$ becomes negligible, and the variance of $\hat{\mu}_j^z$ can be approximated by $\gamma_{\mathcal{C}-\mathcal{B}}$. The approximate standard error of $\hat{\mu}_j^z$ is $\sqrt{\hat{\gamma}_{\mathcal{C}-\mathcal{B}}}$, where $\hat{\gamma}_{\mathcal{C}-\mathcal{B}}$ is the estimate of $\gamma_{\mathcal{C}-\mathcal{B}}$. It is clear that $\gamma_{\mathcal{C}-\mathcal{B}}$ is larger than $\gamma_{\mathcal{B}-\mathcal{A}}$ and $\gamma_{\mathcal{A}}$, which implies that the calibrated estimate $\hat{\mu}_j^z$ based on the RNA-Seq measurement alone is less accurate than $\hat{\mu}_j^{yz}$ and $\hat{\mu}_j^{xyz}$. Because $\gamma_{\mathcal{C}-\mathcal{B}}$ is equal to σ_3^2/β_3^2 , which is the reproducibility of Y_j , $\hat{\mu}_j^z$ has the same reproducibility as the RNA-Seq measurement Y_j . The reason for using $\hat{\mu}_j^z$ is that it is in the same scale as $\hat{\mu}_j^{xyz}$ and $\hat{\mu}_j^{yz}$ so that the calibrated estimates in $\mathcal{C} - \mathcal{B}$ are comparable with those in \mathcal{A} and $\mathcal{B} - \mathcal{A}$.

The calibrated estimates $\{\hat{\mu}_j^{xyz} : j \in \mathcal{A}\}$, $\{\hat{\mu}_j^{yz} : j \in \mathcal{B} - \mathcal{A}\}$ and $\{\hat{\mu}_j^z : j \in \mathcal{C} - \mathcal{B}\}$ can also be considered gene expression measurements normalized by the qRT-PCR platform. As discussed above, the standard errors of these three types of calibrated estimates are different, which are $\sqrt{\hat{\gamma}_{\mathcal{A}}}$, $\sqrt{\hat{\gamma}_{\mathcal{B}-\mathcal{A}}}$, and $\sqrt{\hat{\gamma}_{\mathcal{C}-\mathcal{B}}}$ for genes in \mathcal{A} , $\mathcal{B} - \mathcal{A}$ and $\mathcal{C} - \mathcal{B}$, respectively. When used for gene expression analysis such as detecting differentially expressed genes, both the calibrated estimates and their standard errors need to be used.

2.2. Multi-lab Scenario

Under the multi-lab scenario, it is possible to decompose the measurement error due to a platform into the measurement error due to the technology of the platform and the measurement error due to lab. For example, the measurement error due to the qRT-PCR platform can be decomposed into the measurement error due to qRT-PCR and the measurement error due to lab. Denote the qRT-PCR measurement data by X_{jk} where j and k are indices for genes and labs,

respectively, with $1 \leq k \leq k_1$. Similarly, denote the microarray measurement data by Y_{jk} with $1 \leq k \leq k_2$ and the RNA-Seq measurement data by Z_{jk} with $1 \leq k \leq k_3$. The involved labs are considered to be nested within the platforms. For convenience, in the rest of this article, we refer to the qRT-PCR, microarray and RNA-Seq platforms as the 1st, 2nd and 3rd platforms, respectively. The following system of ME models can be used to model the gene expression measurements obtained by the three platforms from multiple labs,

$$X_{jk} = \mu_j + \kappa_{1j} + \pi_{1jk} \quad (1 \leq j \leq n; 1 \leq k \leq k_1), \quad (2.10a)$$

$$Y_{jk} = \alpha_2 + \beta_2 \mu_j + \kappa_{2j} + \pi_{2jk} \quad (1 \leq j \leq m; 1 \leq k \leq k_2), \quad (2.10b)$$

$$Z_{jk} = \alpha_3 + \beta_3 \mu_j + \kappa_{3j} + \pi_{3jk} \quad (1 \leq j \leq l; 1 \leq k \leq k_3), \quad (2.10c)$$

where κ_{ij} is the measurement error due to the technology of the i th platform for $1 \leq i \leq 3$, and π_{ijk} is the measurement error due to the k th lab of the i th platform. We further assume that the measurement errors are independent with each other, κ_{ij} 's are i.i.d. $N(0, \varsigma_i^2)$, and π_{ijk} 's are i.i.d. $N(0, \psi_i^2)$. Note that when $k_1 = k_2 = k_3 = 1$, the system of ME models (2.10) under the multi-lab scenario reduces to the system of ME models (2.1) under the single-lab scenario, and $\epsilon_{ij} = \kappa_{ij} + \pi_{ij1}$ for $1 \leq i \leq 3$. When $k_1, k_2, k_3 \geq 2$, the structural and incidental parameters in the system of ME models (2.10) can be estimated in a similar way as those in the system of ME models (2.1). Due to limited space, these results are not reported in this article, instead, they are included in Appendix D.

The real data set we will analyze in Section 4 contains qRT-PCR data from one lab, RNA-Seq data from two labs, and microarray data from five labs. In what follows, we present the estimation results for the system of ME models (2.10) with $k_1 = 1$, $k_2 \geq 2$ and $k_3 \geq 2$. When $k_1 = 1$, κ_{1j} and π_{1j1} become inseparable, and consequently, the two variance components ς_1^2 and ψ_1^2 cannot be estimated separately, instead, only their sum $\sigma_1^2 = \varsigma_1^2 + \psi_1^2$ is estimable.

Since $k_1 = 1$, we suppress the index k in X_{jk} and define \bar{X}_j the same way as under the single-lab scenario. Let $\bar{Y}_{j\cdot}$ be the mean of Y_{jk} over the k_2 labs, and \bar{Y}_\cdot is the grand mean across all of the genes in \mathcal{A} . For the RNA-Seq data, $\bar{Z}_{j\cdot}$, and \bar{Z}_\cdot are defined in similar ways as $\bar{Y}_{j\cdot}$, and \bar{Y}_\cdot of the microarray data. Denote the sample variances and covariances of $\{X_j\}_{1 \leq j \leq n}$, $\{\bar{Y}_{j\cdot}\}_{1 \leq j \leq n}$ and $\{\bar{Z}_{j\cdot}\}_{1 \leq j \leq n}$ as $S_{xx}, S_{yy}, S_{zz}, S_{xy}, S_{xz}$ and S_{yz} . The two-step approach used for parameter estimation under the single-lab scenario can still be used under the multi-lab scenario. First, we derive the estimates of the structural parameters. The estimates of the linear coefficients α_i and β_i for $i = 2, 3$ are $\hat{\beta}_2 = S_{yz}/S_{xz}$, $\hat{\beta}_3 = S_{yz}/S_{xy}$, $\hat{\alpha}_2 = \bar{Y}_\cdot - \hat{\beta}_2 \bar{X}_\cdot$, and $\hat{\alpha}_3 = \bar{Z}_\cdot - \hat{\beta}_3 \bar{X}_\cdot$. The estimates of the other structural parameters are given as follows.

$$\hat{\sigma}_1^2 = \left(S_{xx} - \frac{S_{xy}S_{xz}}{S_{yz}} \right), \quad (2.11a)$$

$$\hat{\psi}_2^2 = \frac{1}{nk_2} \sum_{j,k} (Y_{jk} - \bar{Y}_{j\cdot})^2, \quad \hat{\psi}_3^2 = \frac{1}{nk_3} \sum_{j,k} (Z_{jk} - \bar{Z}_{j\cdot})^2, \quad (2.11b)$$

$$\hat{\varsigma}_2^2 = \left(S_{yy} - \frac{S_{xy}S_{yz}}{S_{xz}} \right) - \frac{\hat{\psi}_2^2}{k_2}, \quad \hat{\varsigma}_3^2 = \left(S_{zz} - \frac{S_{yz}S_{xz}}{S_{xy}} \right) - \frac{\hat{\psi}_3^2}{k_3}. \quad (2.11c)$$

Let ϕ be the vector of the structural parameters, that is, $\phi = (\alpha_2, \alpha_3, \beta_2, \beta_3, \sigma_1^2, \psi_2^2, \psi_3^2, \varsigma_2^2, \varsigma_3^2)^\top$

, and $\hat{\phi}$ the vector of the estimated structural parameters. The following proposition establishes the asymptotic properties of $\hat{\phi}$ under the functional system of ME models with multiple labs as n goes to ∞ .

Proposition 5. *Assume the system of ME models (2.10) with $k_1 = 1$, $k_2 \geq 2$ and $k_3 \geq 2$ holds, and the limits in (2.3) exist. Then, as n goes to ∞ , $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(\mathbf{0}, \Lambda_\phi)$, where \xrightarrow{D} means converge in distribution, and Λ_ϕ is the variance-covariance matrix with its explicit expression given in Appendix E.*

Once the estimates of the structural parameters are obtained, we can plug them back into the system of ME models so that the incidental parameters (i.e. μ_j 's) can be estimated by the generalized least squares method. Again, the resulting estimate of μ_j depends on whether j is in \mathcal{A} , $\mathcal{B} - \mathcal{A}$, or $\mathcal{C} - \mathcal{B}$ as under the single-lab scenario. Define $\lambda_1 = \sigma_1^2$, and $\lambda_i = \varsigma_i^2 + \psi_i^2/k_i$ for $i = 2, 3$. Let $\hat{\lambda}_i$ be the estimate of λ_i for $1 \leq i \leq 3$. The estimates of the incidental parameters are given as follows.

$$\hat{\mu}_j^{xyz} = \frac{X_j/\hat{\lambda}_1 + \hat{\beta}_2(\bar{Y}_j - \hat{\alpha}_2)/\hat{\lambda}_2 + \hat{\beta}_3(\bar{Z}_j - \hat{\alpha}_3)/\hat{\lambda}_3}{1/\hat{\lambda}_1 + \hat{\beta}_2^2/\hat{\lambda}_2 + \hat{\beta}_3^2/\hat{\lambda}_3} \quad \text{for } j \in \mathcal{A}, \quad (2.12a)$$

$$\hat{\mu}_j^{yz} = \frac{\hat{\beta}_2(\bar{Y}_j - \hat{\alpha}_2)/\hat{\lambda}_2 + \hat{\beta}_3(\bar{Z}_j - \hat{\alpha}_3)/\hat{\lambda}_3}{\hat{\beta}_2^2/\hat{\lambda}_2 + \hat{\beta}_3^2/\hat{\lambda}_3} \quad \text{for } j \in \mathcal{B} - \mathcal{A}, \quad (2.12b)$$

$$\hat{\mu}_j^z = \frac{\bar{Z}_j - \hat{\alpha}_3}{\hat{\beta}_3} \quad \text{for } j \in \mathcal{C} - \mathcal{B}. \quad (2.12c)$$

The variances of the three types of estimates given above have the same forms as their counterparts under the single-lab scenario. The only difference is that σ_i^2 under the single-lab scenario should be replaced by λ_i under the multi-lab scenario. Based on their asymptotic variances, the standard errors of the estimates can also be obtained. The estimates $\{\hat{\mu}_j^{xyz} : j \in \mathcal{A}\}$, $\{\hat{\mu}_j^{yz} : j \in \mathcal{B} - \mathcal{A}\}$ and $\{\hat{\mu}_j^z : j \in \mathcal{C} - \mathcal{B}\}$, together with their standard errors, can be used in downstream gene expression analysis.

3. Simulation

An extensive simulation study was conducted to evaluate the performance of the proposed calibration methods. In this section, we report the simulation study results for three settings of the system of ME models under the single-lab scenario. Simulation study results under the multi-lab scenario are similar to those under the single-lab scenario, and thus are not reported in this article.

The structural parameters of the three settings are listed in Table 1. The incidental parameters (i.e. μ_j 's) for each model setting were randomly drawn from $N(0, 5)$. For each model setting, we independently generated a training data set of 300 genes and a testing data set of 1000 genes. Both data sets contain the measurements of the genes by all three platforms. The measurements of the first n genes in the training set were used to estimate the structural parameters

(n was varied from 20 to 300), and then the three types of calibrated estimates of the expression levels of the genes in the testing set were obtained. For each combination of model setting and n , this procedure was repeated 200 times to calculate the MSEs of the estimates for the purpose of performance evaluation and comparison.

TABLE 1
The Settings of Model Parameters in Simulation Study

	Intercept		slope		Reproducibility		
	α_2	α_3	β_2	β_3	τ_1^2	τ_2^2	τ_3^2
Setting 1	-0.5	0.1	1	1	0.6	1.5	0.8
Setting 2	0.02	0.2	0.9	0.95	0.5	1	0.75
Setting 3	-0.2	0.01	1.3	1.2	0.6	1	1.2

We first examine and compare the variances of the calibrated expression levels $\hat{\mu}_j^{xyz}$, $\hat{\mu}_j^{yz}$, and $\hat{\mu}_j^z$. Recall that $\text{Var}(\hat{\mu}_j^{xyz}) = \gamma_{\mathcal{A}} + n^{-1}\omega_{\mathcal{A}}(\boldsymbol{\theta}) + O(n^{-2})$. When n is large, $\gamma_{\mathcal{A}}$ is the dominating term, and the other higher order terms are negligible; but when n is moderate (e.g. $n \in [20, 50]$), the second term $n^{-1}\omega_{\mathcal{A}}(\boldsymbol{\theta})$ is not negligible. Further notice that $\gamma_{\mathcal{A}}$ depends on the structural parameters only, but $n^{-1}\omega_{\mathcal{A}}(\boldsymbol{\theta})$ also depends on the incidental parameters, particularly $(\mu_j - \bar{\mu})^2$ as shown in Appendix C. This implies that the variance of $\hat{\mu}_j^{xyz}$ increases as μ_j deviates away from the mean $\bar{\mu}$ when n is moderate. The variances of $\hat{\mu}_j^{yz}$ and $\hat{\mu}_j^z$ behave in the same way as the variance of $\hat{\mu}_j^{xyz}$.

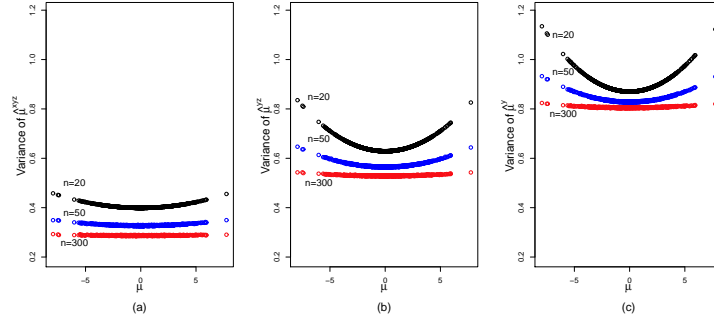


Fig 1: Variances of calibrated estimates of gene expression levels

Plots (a), (b) and (c) are for $\{\text{Var}(\hat{\mu}_j^{xyz})\}$, $\{\text{Var}(\hat{\mu}_j^{yz})\}$ and $\{\text{Var}(\hat{\mu}_j^z)\}$, respectively.

Figure 1 shows the plots of the sample variances of $\hat{\mu}_j^{xyz}$, $\hat{\mu}_j^{yz}$, and $\hat{\mu}_j^z$ versus the incidental parameters μ_j for the second model setting and three different sizes of training data subsets (i.e. $n = 20, 50$, and 300). For convenience, we refer to the curve generated by plotting the variance of a type of estimate against the incidental parameter as the *variance curve* of the type of estimate. Plot (a) in the left panel includes the variance curves of $\hat{\mu}_j^{xyz}$ for $n = 20, 50$, and 300 from the top to the bottom, respectively, Plot (b) in the middle includes those of $\hat{\mu}_j^{yz}$,

and Plot (c) in the right panel includes those of $\hat{\mu}_j^z$. In each plot, the variance curve for $n = 20$ is above that for $n = 50$, which is above that for $n = 300$, indicating that as n increases, the variance of the respective estimate decreases. Furthermore, in each plot, the variance curve for $n = 20$ demonstrates a strong quadratic pattern, indicating its dependence on $(\mu_j - \bar{\mu})^2$, the variance curve for $n = 50$ shows a much mitigated quadratic pattern, and the variance curve for $n = 300$ becomes flat. Comparing the variance curves across the plots, it is clear that under the same n , the variance curve of $\hat{\mu}_j^{xyz}$ is lower than that of $\hat{\mu}_j^{yz}$, which is lower than that of $\hat{\mu}_j^z$, indicating that in terms of accuracy, the order of the three types of estimates, from the best to the worst, is $\hat{\mu}_j^{xyz}$, $\hat{\mu}_j^{yz}$ and $\hat{\mu}_j^z$. These results confirm the properties of the variances of the three types of estimates discussed in Section 2.

We further examine the average MSEs of the three types of calibrated estimates over all of the genes in the testing data set. For convenience in discussion, we refer to the scatter plot of the average MSE of a type of estimate over all of the genes in the testing data set versus the training data size n as the *aMSE curve* of the type of estimate. Note that X_j is also an unbiased estimate of μ_j , we also compare the average MSEs of X_j with the three types of estimates. Figure 2 demonstrates the aMSE curves of the three types of estimates $\{\hat{\mu}_j^{xyz}\}$, $\{\hat{\mu}_j^{yz}\}$, and $\{\hat{\mu}_j^z\}$ as well as that of the original qRT-PCR measurements $\{\bar{X}_j\}$. Plots (a), (b) and (c) in Figure 2 from the left panel to the right panel correspond to the three model settings given in Table 1, respectively.

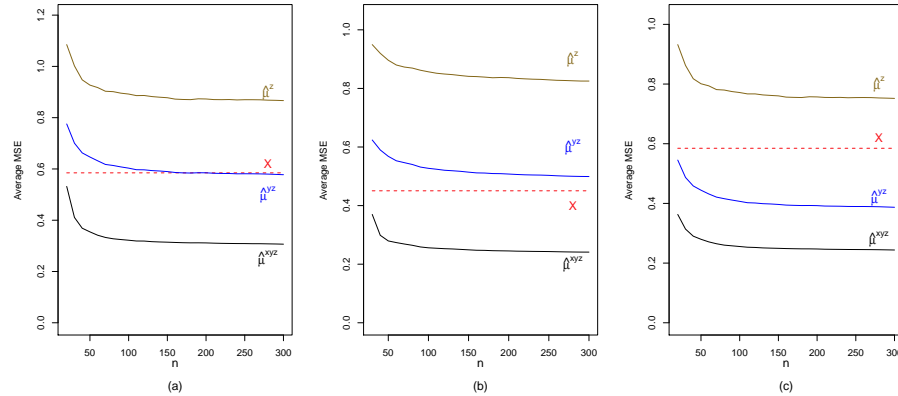


Fig 2: Average MSEs of Original Measurements and Calibrated Estimates

Plots (a), (b) and (c) are for model settings 1, 2 and 3, respectively.

The aMSE curves of $\{X_j\}$ is flat because they do not depend on n . Overall, the aMSE curves of $\{\hat{\mu}_j^{xyz}\}$, $\{\hat{\mu}_j^{yz}\}$, and $\{\hat{\mu}_j^z\}$ decrease as n increases, and they become flat after n is greater than 100, indicating that the structural parameters are accurately estimated and the benefit of further increasing n in the training data set becomes negligible.

In all of the plots, the aMSE curves of $\{\hat{\mu}_j^{xyz}\}$ are always the lowest, indicating that $\{\hat{\mu}_j^{xyz}\}$ give the most accurate quantification of the gene expression levels. The aMSE curves of $\{\hat{\mu}_j^{yz}\}$ are always below the aMSE curves of $\{\hat{\mu}_j^z\}$, indicating that $\{\hat{\mu}_j^{yz}\}$ is more accurate than $\{\hat{\mu}_j^z\}$. However, the aMSE curve of $\{\hat{\mu}_j^{yz}\}$ is not always below that of $\{X_j\}$, indicating that $\{\hat{\mu}_j^{yz}\}$ are not necessarily more accurate than the qRT-PCR measurements. In Plot (a), when n is sufficiently large (e.g. ≥ 100), the aMSE curve of $\{\hat{\mu}_j^{yz}\}$ nearly overlaps with that of $\{X_j\}$, indicating that under the first model setting and when the estimates of the structural parameters are sufficiently accurate, the calibrated expression levels are as accurate as the qRT-PCR measurements $\{X_j\}$. In Plot (b), the aMSE curve of $\{\hat{\mu}_j^{yz}\}$ is above that of $\{X_j\}$, indicating that for the second model setting, $\{\hat{\mu}_j^{yz}\}$ successfully calibrated the RNA-Seq and microarray measurements towards the qRT-PCR measurements but did not outperform qRT-PCR. In Plot (c), the aMSE curve of $\{\hat{\mu}_j^{yz}\}$ is below that of $\{X_j\}$, indicating that $\{\hat{\mu}_j^{yz}\}$ are more accurate than $\{X_j\}$ under the third modeling setting. The performance of $\{\hat{\mu}_j^{yz}\}$ relative to $\{X_j\}$ depends on the model settings. In all of the plots, the aMSE curves of $\{\hat{\mu}_j^z\}$ is always above that of $\{X_j\}$, indicating that the calibrated estimate based on RNA-Seq measurement alone cannot outperform the qRT-PCR measurement.

As discussed in Section 2, the structural parameters are fundamentally different from the incidental parameters. As n increases, the estimates of the structural parameters will converge to their respective targets as shown in Proposition 1. The simulation results about the structural parameters can be found in Appendix F.

4. Application and Results

4.1. Data Sets

We applied the system of ME models to analyze the qRT-PCR, RNA-Seq, and microarray gene expression data of two RNA samples, namely the human brain reference RNA sample (or in short, the Brain sample) and the human universal reference RNA sample (or the UHR sample), generated by the Microarray Quality Control (MAQC) and Sequencing Quality Control (SEQC) projects.

The qRT-PCR data were generated by one single lab using TaqMan® Gene Expression Assays and can be downloaded from Gene Expression Omnibus (GEO) with the series number GSE5350 (<ftp://ftp.ncbi.nih.gov/pub/geo/DATA/supplementary/series/GSE5350/>). The data originally contain the qRT-PCR measurements of 1001 genes, among them 7 genes have multiple entries with distinct expression values but under the same RefSeq ID. To avoid ambiguity, these genes were removed. Each gene has 4 technical replicates for each of the UHR and Brain samples.

The RNA-Seq data include measurements generated from two RNA-Seq experiments conducted in two different labs using Illumina Genome Analyzer. The data from the first lab [5] can be downloaded from NCBI Sequence Read Archive

(SRA) (<http://www.ncbi.nlm.nih.gov/sra/>) under the accession number SRA010153, and the data from the second lab can be downloaded from the same website under the accession number SRA008403. For convenience, we will use the accession numbers to refer to these two data in the rest of the article. The RPKM value is calculated lane-by-lane for each gene. For each of the Brain and UHR sample, those genes that received no reads in at least one lane of the two RNA-Seq experiments were excluded. (See Appendix G for more information of the two RNA-Seq data.)

The microarray data were generated by five different microarray experiments conducted in five different labs using Affymetrix U133 Plus2.0. For convenience, we label the labs as MA1 to MA5 in the rest of the article. The original probe level data can be downloaded from Gene Expression Omnibus (GEO) with the series number GSE5350 (<ftp://ftp.ncbi.nih.gov/pub/geo/DATA/supplementary/series/GSE5350/>). Data from each lab have 5 replicates for both UHR and Brain samples. For each replicate in each lab, the average probe-level measurement of a gene is considered the gene's expression level intensity.

We integrated the three different types of gene expression data as follows. First, the log-2 transformation was applied to all three types of gene expression data. Then for each gene and each lab, the mean measurement across technical replicates is used as the gene's expression value. Furthermore, we exclude genes with extremely low or extremely high expression levels (i.e. those with qRT-PCR expression measurements below -6 or above 4 in log-2 scale) in \mathcal{A} , the remaining genes in \mathcal{A} are used to estimate the structural parameters. For the Brain sample, there are 409 genes that have expression data from all three platforms (\mathcal{A}), 6419 genes that have only RNA-Seq and microarray measurements ($\mathcal{B} - \mathcal{A}$), and 5949 genes that only have RNA-Seq measurements ($\mathcal{C} - \mathcal{B}$); for the UHR sample, there are 477 genes that have measurements by all three platforms (\mathcal{A}), 7187 genes that have measurements only by RNA-Seq and microarray ($\mathcal{B} - \mathcal{A}$), and 6892 genes that have only measurements by RNA-Seq ($\mathcal{C} - \mathcal{B}$).

Two schemes were used to analyze the integrated data. First, we considered the single-lab scenario, in which each platform has data from one lab. In total there are ten possible combinations, and we applied the system of ME models (2.1a)-(2.1c) to each combination. The analysis results of individual combinations were similar, and we only report the results for the combination that includes the RNA-Seq data from SRA010153 and the microarray data from MA1. Second, we applied the system of ME models for the multi-lab scenario to analyze the entire data set for each RNA sample. Due to limited space, the results from the multi-lab scenario are presented and briefly discussed in Appendix I.

4.2. Results under Single-Lab Scenario

4.2.1. Diagnostics of Model Assumptions

The system of ME models (2.1a)-(2.1c) imposes the normality and homoscedasticity assumptions on the measurement errors. We checked these assumptions

using genes in \mathcal{A} . After the system of ME models was fitted, we calculated the residuals by $e_{1j} = X_j - \hat{\mu}_j$, $e_{2j} = Y_j - \hat{\alpha}_2 - \hat{\beta}_2 \hat{\mu}_j$ and $e_{3j} = Z_j - \hat{\alpha}_3 - \hat{\beta}_3 \hat{\mu}_j$ corresponding to the measurement errors due to the qRT-PCR, microarray and RNA-Seq platforms, respectively. To check the normality assumption, we generated the QQ plots for the residuals, and did not detect significant violation of the assumption. To check the homoscedasticity assumption, we generated residual plots and constructed approximate 95% confidence intervals of the Box-Cox transformation. Because the diagnostic results are similar for the two RNA samples, we only present those from the UHR sample as an example. The residual plots corresponding to the qRT-PCR, microarray, and RNA-Seq platforms are presented in Figure 7 in the Appendix H. The plots do not demonstrate strong heteroscedastic patterns. The 95% confidence intervals of Box-Cox transformation for the residuals from the three platforms are presented in Figure 8 in Appendix H. The confidence intervals corresponding to the qRT-PCR and RNA-Seq platforms both contain 1, and 1 is on the boundary of the confidence interval corresponding to the microarray platform. These results together indicate that there does not exist significant violation of the homoscedasticity assumption imposed on the platform measurement errors.

4.2.2. Structural Parameters

The estimates of the structural parameters for the Brain and UHR samples are given in Table 2. The standard errors of the estimates are also reported in parentheses in the table. The estimates of the structural parameters can be used to compare the three platforms in terms of the quality of measurements they provide.

From models (2.1a)-(2.1c), the qRT-PCR measurements are unbiased with respect to μ_j 's, and the intercepts α_2 and α_3 represent the biases of microarray and RNA-Seq measurements relative to the qRT-PCR measurements. Larger absolute values of α_2 and α_3 indicate larger biases. From Table 2, $\hat{\alpha}_2$ are 8.8401 and 9.1033 in the Brain and UHR samples, respectively; and $\hat{\alpha}_3$ are 4.9405 and 5.4249 in the Brain and UHR samples, respectively. In both samples, $\hat{\alpha}_2 > \hat{\alpha}_3$, indicating that microarray measurements have larger biases than the RNA-Seq measurements. The slopes β_2 and β_3 represent the scales of microarray and RNA-Seq measurements relative to qRT-PCR measurements. From Table 2, $\hat{\beta}_2$ are 0.7754 and 0.7695 in the Brain and UHR samples, respectively, both of which are significantly less than 1, indicating that the microarray measurements are in a smaller scale compared to the qRT-PCR measurements. On the other hand, $\hat{\beta}_3$ are 1.0254 and 1.0009 in the Brain and UHR samples, respectively, both of which are not significantly different from 1, indicating that the RNA-Seq measurements are in the similar scale as the qRT-PCR measurements. The variances σ_1^2 , σ_2^2/β_2^2 and σ_3^2/β_3^2 reflect the reproducibilities of the three platforms. Smaller value indicates higher reproducibility. From Table 2, in the Brain sample, the three values are 0.7945, 2.0635 and 1.0157; and in the UHR sample, the three values are 0.6685, 1.7638 and 0.9435. In both samples, $\hat{\sigma}_1^2 < \hat{\sigma}_3^2/\hat{\beta}_3^2 < \hat{\sigma}_2^2/\hat{\beta}_2^2$,

indicating that qRT-PCR has the best reproducibility, microarray has the worst reproducibility, and RNA-Seq is slightly worse than qRT-PCR but much better than microarray.

TABLE 2
Estimates of Structural Parameters Using the qRT-PCR Data, RNA-Seq Data from SRA010153, and microarray Data from MA1

	Intercept		Slope		Variance		
	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$
Brain	8.8401 (0.0724)	4.9405 (0.0859)	0.7754 (0.0216)	1.0254 (0.0324)	0.7945 (0.1143)	1.2407 (0.1031)	1.0679 (0.1283)
UHR	9.1033 (0.0605)	5.4249 (0.0692)	0.7695 (0.0212)	1.0009 (0.0146)	0.6685 (0.0947)	1.0444 (0.0836)	0.9452 (0.1040)

4.2.3. Incidental Parameters

After the estimates of the structural parameters were obtained, we used the formulas obtained in Section 2 to estimate the gene expression levels, and the standard errors of the calibrated gene expression values were also calculated. These calibrated gene expression values together with their standard errors are expected to have a higher quality than the original measurements by the three platforms and lead to better results in downstream gene expression analysis.

We compare the calibrated gene expression levels with their original qRT-PCR, microarray, and RNA-Seq measurements, and use the genes in the Brain and UHR samples that have measurements by all three platforms as an illustrative example. Because these genes have measurements by all three platforms, all three types of calibrated estimates $\{\hat{\mu}_j^z\}$, $\{\hat{\mu}_j^{yz}\}$, and $\{\hat{\mu}_j^{xyz}\}$ could be calculated. Based on the theoretical and simulation results in Sections 2 and 3, $\{\hat{\mu}_j^{xyz}\}$ are the most accurate measurements. We plotted the original measurements $\{X_j\}$, $\{Y_j\}$ and $\{Z_j\}$ as well as the calibrated measurements $\{\hat{\mu}_j^{yz}\}$ against $\{\hat{\mu}_j^{xyz}\}$, and presented the plots in Figure 3 and Figure 4 for the Brain and UHR sample, respectively. Because $\{\hat{\mu}_j^z\}$ are the linear transformation of the original RNA-Seq measurements $\{Z_j\}$, the plot of $\{\hat{\mu}_j^z\}$ versus $\{\hat{\mu}_j^{xyz}\}$ had the same appearance as the plot of $\{Z_j\}$ versus $\{\hat{\mu}_j^{xyz}\}$ and thus was not presented. In each plot, the correlation coefficient between the two plotted measurements was calculated and reported.

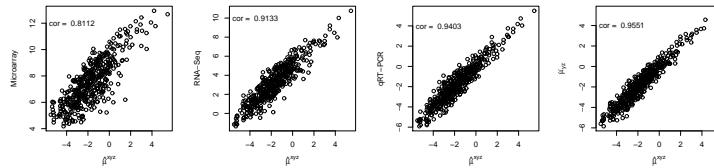


Fig 3: Scatter Plots of the measurements by microarray, RNA-Seq, qRT-PCR and $\hat{\mu}^{yz}$ versus $\hat{\mu}^{xyz}$ on \mathcal{A} in the Brain Sample

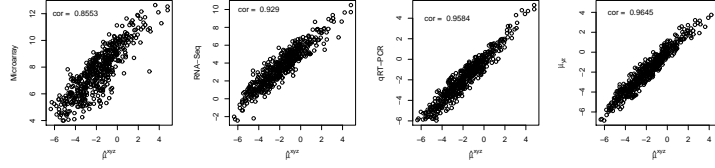


Fig 4: Scatter Plots of the measurements by microarray, RNA-Seq, qRT-PCR and $\hat{\mu}^{yz}$ versus $\hat{\mu}^{xyz}$ on \mathcal{A} in the UHR Sample

From Figure 3, in the Brain sample, in terms of the linear correlation coefficient with the best measurements $\{\hat{\mu}_j^{xyz}\}$, microarray is the worst (0.8112) and qRT-PCR is the best (0.9403) among the three platforms. This is consistent with the findings reported in the literature that in terms of the quality of the gene expression data produced by the three platforms, qRT-PCR is most accurate, and microarray is least accurate. The calibrated expression levels $\{\hat{\mu}_j^{yz}\}$ have a higher correlation coefficient with $\{\hat{\mu}_j^{xyz}\}$ (0.9551) than the qRT-PCR, microarray, and RNA-Seq measurements. In the UHR sample, in terms of the linear correlation coefficient with the best measurements $\{\hat{\mu}_j^{xyz}\}$, again microarray is the worst (0.8553) and qRT-PCR is the best (0.9583) among the three platforms. The calibrated expression levels $\{\hat{\mu}_j^{yz}\}$ have a higher correlation coefficient with $\{\hat{\mu}_j^{xyz}\}$ (0.9645) than qRT-PCR, microarray, and RNA-Seq measurements.

4.2.4. Gene Differential Expression

The goal of gene differential expression analysis is to identify genes that are differentially expressed across different biological or environmental conditions. It is of interest to find genes that have different expression levels in the Brain and UHR samples. To demonstrate that calibrated expression levels can lead to better gene differential expression analysis, we used the RNA-Seq measurements (Z_j 's) and the calibrated measurements ($\hat{\mu}_j$'s), separately, to perform gene differential expression analysis for the two RNA samples. For both the calibrated measurements and the RNA-Seq measurements, we used two methods, which are z -test and local FDR [9], to detect differentially expressed genes and compared their results. The results from these two methods are discussed separately below.

For each RNA sample, the calibrated expression level of a gene was $\hat{\mu}_j^{xyz}$, $\hat{\mu}_j^{yz}$, or $\hat{\mu}_j^z$ depending on whether the gene is in \mathcal{A} , $\mathcal{B} - \mathcal{A}$, or $\mathcal{C} - \mathcal{B}$. In what follows, for convenience, we do not further distinguish the notation for these three types of calibrated estimates. Instead, we use $\hat{\mu}_j^{(UHR)}$ and $\hat{\mu}_j^{(Brain)}$ to denote the calibrated expression levels of gene j in the UHR and Brain samples, respectively, and use $\hat{\varphi}_j^{(UHR)}$ and $\hat{\varphi}_j^{(Brain)}$ to denote the estimated variances of $\hat{\mu}_j^{(UHR)}$ and $\hat{\mu}_j^{(Brain)}$, respectively.

When performing z -test based on the calibrated measurements for gene j , the p -value for testing $H_0 : \mu_j^{(UHR)} = \mu_j^{(Brain)}$ was

$$p_j = 2P \left(Z > \frac{|\hat{\mu}_j^{(Brain)} - \hat{\mu}_j^{(UHR)}|}{\sqrt{\hat{\varphi}_j^{(Brain)} + \hat{\varphi}_j^{(UHR)}}} \right) \quad (4.1)$$

where Z follows the standard normal distribution. Based on the calculated p -values, the standard Benjamini-Hochberg procedure ([4]) was used to identify differentially expressed genes with the false discovery rate (FDR) controlled at 0.01. The numbers of the identified genes are reported in Table 3. Similarly, based on the RNA-Seq measurements, we applied z -test together with the Benjamini-Hochberg procedure to detect differentially expressed genes between the UHR and Brain samples. The numbers of the identified differentially expressed genes based on the RNA-Seq measurements are reported in the second column of Table 3.

From Table 3, in total, 407 genes were detected to be differentially expressed in the UHR and Brain samples using the calibrated expression measurements, whereas only 158 genes were detected using the original RNA-Seq measurements. These two groups of genes share 154 genes. Therefore, the calibrated expression measurements led to the detection of almost all the genes (154 out of 158 genes) detected by the original RNA-Seq measurements. In addition, the former detected 249 more genes than the latter. The total number of genes detected by both types of measurements was further broken down according to whether a gene belongs to \mathcal{A} , $\mathcal{B} - \mathcal{A}$, or $\mathcal{C} - \mathcal{B}$, and the results are also reported in Table 3. From the table, 47, 142 and 60 more genes were detected by the calibrated expression measurements in \mathcal{A} , $\mathcal{B} - \mathcal{A}$ and $\mathcal{C} - \mathcal{B}$, respectively, than by the original RNA-Seq measurements. Therefore, at the same FDR, the calibrated measurements led to a higher power for gene differential expression analysis than the original RNA-Seq measurements when z -test was used.

In examining the selected genes, the 253 genes that are selected by the calibrated expression levels but not by the RNA-Seq measurements are of particular interest. The annotations of these genes suggest that half of them are related to brain functions or have been involved in past brain related studies. Among these genes, some are found to be brain specific genes such as *FABP7*, *KCNA6*, *SNPH*, *CREG2* and ; some are found to be highly expressed in brain, for example *TRIL*, *CYP26B1*, *RAPGEF5*, *EHD3* and *PPP2R5B*; and some are found to be lowly expressed in brain, for example *PAQR6*.

TABLE 3
Numbers of Differentially Expressed Genes Detected by Calibrated Estimates and RNA-Seq Measurements

Gene Set	z-test			Local FDR		
	Calibration	RNA-Seq	Overlap	Calibration	RNA-Seq	Overlap
\mathcal{A}	56	9	9	13	8	8
$\mathcal{B} - \mathcal{A}$	202	60	60	153	164	135
$\mathcal{C} - \mathcal{B}$	149	89	85	220	176	176
Total	407	158	154	386	348	319

For both the calibrated and RNA-Seq measurements of the two RNA samples, we further used the local FDR method to identify differentially expressed genes with the false discovery rate controlled at 0.01. The numbers of the identified differentially expressed genes are reported in the fourth and fifth columns of Table 3.

From Table 3, in total, 386 genes were detected to be differentially expressed in the UHR and Brain samples by the calibrated expression measurements, whereas 348 genes were detected by the original RNA-Seq measurements. These two groups of genes share 319 genes. The calibrated measurements led to the detection of 38 more genes than the RNA-Seq measurements. The total number of genes detected by both types of measurements was broken down according to whether a gene belongs to \mathcal{A} , $\mathcal{B} - \mathcal{A}$, or $\mathcal{C} - \mathcal{B}$, and the results are also reported in Table 3. From the table, 5 and 44 more genes were detected by the calibrated expression measurements in \mathcal{A} and $\mathcal{C} - \mathcal{B}$. In $\mathcal{B} - \mathcal{A}$, the RNA-Seq measurements led to the detection of 11 more genes than the calibrated expression measurements.

Examining the 67 genes that were detected by the calibrated measurements but not by the RNA-Seq measurements, gene annotation suggested that 50 of them are related to brain functions or have been involved in past brain related studies. Among the 29 genes that were detected by the RNA-Seq measurements but not by the calibrated measurements, 20 of them are related to brain functions or have been involved in past brain related studies.

5. Discussion

A system of functional measurement error (ME) models was proposed to calibrate the microarray and RNA-Seq measurements of gene expression levels by qRT-PCR. Due to limited space, the design issue of the proposed approach was not discussed in this article. The success of the proposed approach hinges on the genes that are measured by all three platforms, and the major bottleneck is the relative low throughput of the qRT-PCR platform. Therefore, the design issue is centered on the qRT-PCR platform with respect to two questions. The first question is how many genes should to be measured by qRT-PCR, and the second question is which genes should be measured. For the first question, based on the theoretical, simulation, and real data application results in this article, it seems that at least 150 genes are needed to ensure that the bias and variances

(i.e. the structural parameters) can be accurately estimated and the calibrated gene expression levels (i.e. the incidental parameters) can reach their best possible accuracy. Vendors of the qRT-PCR platform such as Life Technologies now offers assays and services to measure a sufficiently large number of genes simultaneously. This makes our proposed approach feasible in practice. For the second question, based on our theoretical results, the true expression levels of the genes selected to be measured by qRT-PCR should be as spread out as possible. These two questions and the design issue in general will be addressed in a future publication.

As discussed in the introduction, RNA-Seq data are also found to be subject to excessive variability and various methods have been proposed to normalize RNA-Seq data. In this article, only the RPKM method and the resulting RPKM measurements were used as the RNA-Seq measurements. Clearly, other normalization methods and their resulting measurements can also be considered. The variance component due to the RNA-Seq platform estimated under the multi-lab scenario (i.e. ς_2^2) in this article corresponds to the RPKM method. If another normalization method (e.g. the "mseq" method) is used, then ς_2^2 will correspond to that normalization method. Therefore, the system of ME models can be used to compare different normalization methods.

The functional system of ME models of order 3 can be extended to that of order $p > 3$, and the two-step approach to parameter estimation can also be extended in a straightforward manner. In this article, the bias of the measurement of a platform is assumed be a linear function of the true expression level. In general, nonlinear models or even nonparametric models can be considered. Furthermore, covariates can also be incorporated into the system of functional ME models. These possible extensions of the proposed approach will be investigated in the near future.

References

- [1] AFFYMETRIX INC., (2005). Technical Note: Guide to Probe Logarithmic Intensity Error (PLIER) Estimation Affymetrix White Paper.
- [2] APPLIED BIOSYSTEMS, (2006). TaqMan® Gene Expression Assays for Validating Hits From Fluorescent Microarrays White Paper.
- [3] BARNETT, V. D. (1969). Simultaneous Pairwise Linear Structural Relationships. *Biometrics* **25** 129-142.
- [4] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289-300.
- [5] BULLARD, J., PURDOM, E., HANSEN, K. and DUDOIT, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11** 94.
- [6] BUSTIN, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29** 23-39.

- [7] BUSTIN, S. A. and NOLAN, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques* **15** 155–166.
- [8] CHENG, C.-L. and VAN NESS, J. W. (1999). *Statistical Regression with Measurement Error (Kendall's Library of Statistics, 6)*. London: Hodder Arnold.
- [9] EFRON, B. (2007). Size, power and false discovery rates. *The Annals of Statistics* **35** 1351–1377.
- [10] FULLER, W. A. (1987). *Measurement Error Models*. New York: John Wiley & Sons.
- [11] GLESER, L. J. (1983). Functional, Structural and Ultrastructural Errors-in-Variables Models In *Proceedings of the Business and Economic Statistics Section* 57-66. American Statistical Association.
- [12] HU, M., ZHU, Y., TAYLOR, J. M. G., LIU, J. S. and QIN, Z. (2011). Using Poisson mixed-effects model to quantify transcript-level gene expression in RNA-Seq. *Bioinformatics* **28** 63-68.
- [13] KENDALL, M. G. and STUART, A. (1973). *The Advanced Theory of Statistics: Inference and Relationship* **2**, 4th ed. London: Griffin.
- [14] LI, J., JIANG, H. and WONG, W. (2010). Modeling non-uniformity in short-read rates in RNA-Seq data. *Genome Biology* **11** R50.
- [15] MAK, H. C. (2011). John Storey provides his take on the importance of new statistical methods for high-throughput sequencing. *Nature Biotechnology* **29** 331–333.
- [16] MORTAZAVI, A., WILLIAMS, B., MCCUE, K., SCHAEFFER, L. and WOLD, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5** 621-628.
- [17] OSBORNE, C. (1991). Statistical Calibration: A Review. *International Statistical Review / Revue Internationale de Statistique* **59** 309-336.
- [18] PFAFFL, M. W. (2004). Quantification strategies in real-time PCR In *A-Z of Quantitative PCR*. Edited by Bustin S.A. International University Line (IUL), La Jolla.
- [19] REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18** 375-389.
- [20] SCHENA, M., SHALON, D., DAVIS, R. W. and BROWN, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270** 467–470.
- [21] SCHRAGA, S., OREN, R. and AST, G. (2011). Detection and Removal of Biases in the Analysis of Next-Generation Sequencing Reads. *PLoS ONE* **6** e16685.
- [22] SOLARI, M. E. (1969). The "Maximum Likelihood Solution" of the Problem of Estimating a Linear Functional Relationship. *Journal of the Royal Statistical Society. Series B (Methodological)* **31** 372-375.
- [23] SRIVASTAVA, S. and CHEN, L. (2010). A two-parameter generalized Poisson model to improve the analysis of RNA-Seq data. *Nucleic Acids Research* **38** e170.
- [24] WANG, Z., GERSTEIN, M. and SNYDER, M. (2009). RNA-Seq: a revolu-

tionary tool for transcriptomics. *Nat Rev Genet* **10** 57–63.

Appendix A: Proof of Proposition 1

A.1. Asymptotic Consistency of $\hat{\theta}$ under Single-Lab Scenario

Recall that $\Delta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\mu_j - \bar{\mu})^2$. Note that X_j , Y_j and Z_j are independent normal random variables for all j . From classic asymptotic theory, sample moments of normal populations converge to their population counterparts almost surely. Therefore, as $n \rightarrow \infty$, we have

$$\begin{aligned} \bar{X} &\xrightarrow{a.s.} \bar{\mu}, & \bar{Y} &\xrightarrow{a.s.} \alpha_2 + \beta_2 \bar{\mu}, & \bar{Z} &\xrightarrow{a.s.} \alpha_3 + \beta_3 \bar{\mu}, \\ S_{xx} &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \xrightarrow{a.s.} \Delta + \sigma_1^2, \\ S_{yy} &= \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2 \xrightarrow{a.s.} \beta_2^2 \Delta + \sigma_2^2, \\ S_{zz} &= \frac{1}{n} \sum_{j=1}^n (Z_j - \bar{Z})^2 \xrightarrow{a.s.} \beta_3^2 \Delta + \sigma_3^2, \\ S_{xy} &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y}) \xrightarrow{a.s.} \beta_2 \Delta, \\ S_{xz} &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(Z_j - \bar{Z}) \xrightarrow{a.s.} \beta_3 \Delta, \\ S_{yz} &= \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})(Z_j - \bar{Z}) \xrightarrow{a.s.} \beta_2 \beta_3 \Delta. \end{aligned}$$

Based on the above convergence results, we have

$$\begin{aligned} \hat{\beta}_2 &= \frac{S_{yz}}{S_{xz}} \xrightarrow{p} \beta_2, & \hat{\beta}_3 &= \frac{S_{yz}}{S_{xy}} \xrightarrow{p} \beta_3, \\ \hat{\alpha}_2 &= \bar{Y} - \frac{S_{yz}}{S_{xz}} \bar{X} \xrightarrow{p} \alpha_2, & \hat{\alpha}_3 &= \bar{Z} - \frac{S_{yz}}{S_{xy}} \bar{X} \xrightarrow{p} \alpha_3, \\ \hat{\sigma}_1^2 &\xrightarrow{p} ((\Delta + \sigma_1^2) - \beta_2 \Delta \beta_3 \Delta / \beta_2 \beta_3 \Delta) = \sigma_1^2, \\ \hat{\sigma}_2^2 &\xrightarrow{p} ((\beta_2^2 \Delta + \sigma_2^2) - \beta_2^2 \beta_2 \Delta \beta_3 \Delta / \beta_2 \beta_3 \Delta) = \sigma_2^2, \\ \hat{\sigma}_3^2 &\xrightarrow{p} ((\beta_3^2 \Delta + \sigma_3^2) - \beta_3^2 \beta_2 \Delta \beta_3 \Delta / \beta_2 \beta_3 \Delta) = \sigma_3^2. \end{aligned}$$

A.2. Asymptotic Variances of Estimates of the Structural Parameters under Single-Lab Scenario

Let $\mathbf{m} = (\bar{X}, \bar{Y}, \bar{Z}, S_{xx}, S_{yy}, S_{zz}, S_{xy}, S_{xz}, S_{yz})^\top$ be the vector of sample moments. Then $\hat{\boldsymbol{\theta}}$ is a function of \mathbf{m} denoted as $\hat{\boldsymbol{\theta}}(\mathbf{m})$. Proposition 1 can be derived by the delta method, and the covariance matrix of $\hat{\boldsymbol{\theta}}$ is $\boldsymbol{\Gamma}_\theta = \mathbf{J}^\top \mathbf{V} \mathbf{J}$, where \mathbf{J} is the first order derivative of $\hat{\boldsymbol{\theta}}(\mathbf{m})$ w.r.t. \mathbf{m} evaluated at $E(\mathbf{m})$, and \mathbf{V} is the asymptotic covariance matrix of \mathbf{m} . The elements of \mathbf{V} can be obtained explicitly using the properties of moments of Normal distribution as described in Appendix 1.B of Fuller [10]. Let $\eta_i = \sigma_i^2$ for $i = 1, 2, 3$, and $\eta_{ij} = \eta_i \eta_j$ for $i \neq j$. The explicit form of $\boldsymbol{\Gamma}_\theta$ is described as follows.

$$\boldsymbol{\Gamma}_\theta = \begin{bmatrix} \Gamma_{\alpha\alpha} & \Gamma_{\alpha\beta} & \Gamma_{\alpha\sigma} \\ \Gamma_{\alpha\beta}^\top & \Gamma_{\beta\beta} & \Gamma_{\beta\sigma} \\ \Gamma_{\alpha\sigma}^\top & \Gamma_{\beta\sigma}^\top & \Gamma_{\sigma\sigma} \end{bmatrix},$$

where

$$\begin{aligned} \Gamma_{\beta\beta} &= \frac{1}{\Delta^2} \begin{bmatrix} \frac{\beta_2^2}{\beta_3^2} \eta_{13} + \frac{1}{\beta_3^2} \eta_{23} & \frac{\eta_{23}}{\beta_2 \beta_3} \\ \frac{\eta_{23}}{\beta_2 \beta_3} & \frac{\beta_2^2}{\beta_3^2} \eta_{12} + \frac{1}{\beta_3^2} \eta_{23} \end{bmatrix}, \\ \Gamma_{\alpha\alpha} &= \bar{\mu}^2 \Gamma_{\beta\beta} + \begin{bmatrix} \beta_2^2 \eta_1 + \eta_2 & \beta_2 \beta_3 \eta_1 \\ \beta_2 \beta_3 \eta_1 & \beta_3^2 \eta_1 + \eta_3 \end{bmatrix}, \\ \Gamma_{\alpha\beta} &= -\bar{\mu} \Gamma_{\beta\beta}, \\ \Gamma_{\beta\sigma} &= \frac{1}{\Delta} \begin{bmatrix} \frac{\beta_2}{\beta_3^2} \eta_{13} + \frac{1}{\beta_2 \beta_3^2} \eta_{23} & -\left(\frac{\beta_2^3}{\beta_3^2} \eta_{13} + \frac{\beta_2}{\beta_3^2} \eta_{23} \right) & \beta_2 \eta_{13} - \frac{1}{\beta_2} \eta_{23} \\ \frac{\beta_3}{\beta_2^2} \eta_{12} + \frac{1}{\beta_2^2 \beta_3} \eta_{23} & \beta_3 \eta_{12} - \frac{1}{\beta_3} \eta_{23} & -\left(\frac{\beta_3^2}{\beta_2^2} \eta_{12} + \frac{\beta_3}{\beta_2^2} \eta_{23} \right) \end{bmatrix}, \\ \Gamma_{\alpha\sigma} &= -\bar{\mu} \Gamma_{\beta\sigma}, \\ \Gamma_{\sigma\sigma} &= \text{diag} \left(2\eta_1^2 + \frac{2\sigma_1^4}{r_1^3}, 2\eta_2^2 + \frac{2\sigma_2^4}{r_2^3}, 2\eta_3^2 + \frac{2\sigma_3^4}{r_3^3} \right) \\ &\quad + \begin{bmatrix} \frac{\eta_{12}}{\beta_2^2} + \frac{\eta_{13}}{\beta_3^2} + \frac{\eta_{23}}{\beta_2^2 \beta_3^2} & \eta_{12} - \frac{\beta_2^2}{\beta_3^2} \eta_{13} - \frac{\eta_{23}}{\beta_3^2} & \eta_{13} - \frac{\beta_2^2}{\beta_2^2} \eta_{12} - \frac{\eta_{23}}{\beta_2^2} \\ \eta_{12} - \frac{\beta_2^2}{\beta_3^2} \eta_{13} - \frac{\eta_{23}}{\beta_3^2} & \beta_2^2 \eta_{12} + \frac{\beta_2^4}{\beta_3^2} \eta_{13} + \frac{\beta_2^2}{\beta_3^2} \eta_{23} & \eta_{23} - \beta_3^2 \eta_{12} - \beta_2^2 \eta_{13} \\ \eta_{13} - \frac{\beta_2^2}{\beta_3^2} \eta_{12} - \frac{\eta_{23}}{\beta_2^2} & \eta_{23} - \beta_3^2 \eta_{12} - \beta_2^2 \eta_{13} & \beta_3^2 \eta_{13} + \frac{\beta_3^4}{\beta_2^2} \eta_{12} + \frac{\beta_3^2}{\beta_2^2} \eta_{23} \end{bmatrix}. \end{aligned}$$

Appendix B: Estimation of Incidental Parameters

For each $j \in \mathcal{A}$, given the structural parameters, the measurement error models (1a)-(1c) involving μ_j can be rewritten as a linear model of μ_j , that is,

$$\begin{bmatrix} X_j \\ Y_j - \alpha_2 \\ Z_j - \alpha_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \mu_j + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \end{bmatrix}.$$

Replacing the structural parameters $\boldsymbol{\theta}$ with their estimates and applying the generalized least squares method, the estimate for μ_j is

$$\hat{\mu}_j^{xyz} = \frac{X_j/\hat{\sigma}_1^2 + \hat{\beta}_2(Y_j - \hat{\alpha}_2)/\hat{\sigma}_2^2 + \hat{\beta}_3(Z_j - \hat{\alpha}_3)/\hat{\sigma}_3^2}{1/\hat{\sigma}_1^2 + \hat{\beta}_2^2/\hat{\sigma}_2^2 + \hat{\beta}_3^2/\hat{\sigma}_3^2},$$

For each $j \in \mathcal{B} - \mathcal{A}$, given the structural parameters, the measurement error models involving μ_j can be rewritten as

$$\begin{bmatrix} Y_j - \alpha_2 \\ Z_j - \alpha_3 \end{bmatrix} = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} \mu_j + \begin{bmatrix} \epsilon_{2j} \\ \epsilon_{3j} \end{bmatrix}.$$

Replacing the structural parameters $\boldsymbol{\theta}$ with their estimates and applying the generalized least squares method, the estimate of μ_j is

$$\hat{\mu}_j^{yz} = \frac{\hat{\beta}_2(Y_j - \hat{\alpha}_2)/\hat{\sigma}_2^2 + \hat{\beta}_3(Z_j - \hat{\alpha}_3)/\hat{\sigma}_3^2}{\hat{\beta}_2^2/\hat{\sigma}_2^2 + \hat{\beta}_3^2/\hat{\sigma}_3^2}.$$

For each $j \in \mathcal{C} - \mathcal{B}$, given the structural parameters, the measurement error models involving μ_j can be written as

$$Z_j - \alpha_3 = \beta_3 \mu_j + \epsilon_{3j}.$$

Replacing the structural parameters (α_3 , β_3 and σ_3^2) with their estimates, the estimate of μ_j is

$$\hat{\mu}_j^z = (Z_j - \hat{\alpha}_3)/\hat{\beta}_3.$$

Appendix C: Second Order Terms of $\text{Var}(\mu_j^{xyz})$, $\text{Var}(\mu_j^{yz})$ and $\text{Var}(\mu_j^z)$

$$\begin{aligned} \omega_{\mathcal{A}}(\boldsymbol{\theta}) = & \left(\frac{1}{\eta_1} + \frac{\beta_2^2}{\eta_2} + \frac{\beta_3^2}{\eta_3} \right)^{-2} \left\{ \eta_1 \left(\frac{\beta_2^2}{\eta_2} + \frac{\beta_3^2}{\eta_3} \right)^2 + \frac{\beta_2^2}{\eta_2} + \frac{\beta_3^2}{\eta_3} \right. \\ & + \eta_{13} \left[\frac{1}{\eta_1^3 \beta_3^2} + \frac{\beta_3^4}{\eta_3^3} + \eta_2 \left[\frac{\beta_2^3}{\eta_2^2 \beta_3} + \frac{\beta_2}{\eta_2 \beta_3 \Delta} \right]^2 + \frac{\beta_2^4}{\eta_2^2 \beta_3^2 \Delta^2} (\mu_j - \bar{\mu})^2 \right] \\ & + \eta_{12} \left[\frac{1}{\eta_1^3 \beta_2^2} + \frac{\beta_2^4}{\eta_2^3} + \eta_3 \left(\frac{\beta_3^3}{\eta_3^2 \beta_2} + \frac{\beta_3}{\eta_3 \beta_2 \Delta} \right)^2 + \frac{\beta_3^4}{\eta_3^2 \beta_2^2 \Delta^2} (\mu_j - \bar{\mu})^2 \right] \\ & \left. + \eta_{23} \left[\frac{1}{\eta_1^3 \beta_2^2 \beta_3^2} + \eta_3 \left(\frac{\beta_3^2}{\eta_3^2 \beta_2} + \frac{1}{\eta_3 \beta_2 \Delta} \right)^2 + \eta_2 \left(\frac{\beta_2^2}{\eta_2^2 \beta_3} + \frac{1}{\eta_2 \beta_3 \Delta} \right)^2 + \left(\frac{\beta_3}{\eta_3 \beta_2 \Delta} + \frac{\beta_2}{\eta_2 \beta_3 \Delta} \right)^2 (\mu_j - \bar{\mu})^2 \right] \right\}, \end{aligned}$$

$$\begin{aligned}
\omega_{\mathcal{B}-\mathcal{A}}(\boldsymbol{\theta}) &= \left(\frac{\beta_3^2}{\eta_3} + \frac{\beta_2^2}{\eta_2} \right)^{-2} \left\{ \eta_1 \left(\frac{\beta_3^2}{\eta_3} + \frac{\beta_2^2}{\eta_2} \right)^2 - \left(\frac{\beta_3^2}{\eta_3} + \frac{\beta_2^2}{\eta_2} \right) \right. \\
&\quad + \eta_{13} \left[\frac{\beta_3^4}{\eta_3^3} + \eta_2 \left(\frac{\beta_2^3}{\eta_2^2 \beta_3} + \frac{\beta_2}{\eta_2 \Delta \beta_3} \right)^2 + \frac{\beta_2^4}{\eta_2^2 \beta_3^2 \Delta^2} (\mu_j - \bar{\mu})^2 \right] \\
&\quad + \eta_{12} \left[\frac{\beta_2^4}{\eta_2^3} + \eta_3 \left(\frac{\beta_3^3}{\eta_3^2 \beta_2} + \frac{\beta_3}{\eta_3 \beta_2 \Delta} \right)^2 + \frac{\beta_3^4}{\eta_3^2 \beta_2^2 \Delta^2} (\mu_j - \bar{\mu})^2 \right] \\
&\quad \left. + \eta_{23} \left[\eta_3 \left(\frac{\beta_3^2}{\beta_2 \eta_3^2} + \frac{1}{\eta_3 \beta_2 \Delta} \right)^2 + \eta_2 \left(\frac{\beta_2^2}{\beta_3 \eta_2^2} + \frac{1}{\eta_2 \Delta \beta_3} \right)^2 + \left(\frac{\beta_3}{\eta_3 \beta_2 \Delta} + \frac{\beta_2}{\eta_2 \beta_3 \Delta} \right)^2 (\mu_j - \bar{\mu})^2 \right] \right\}, \\
\omega_{\mathcal{C}-\mathcal{B}}(\boldsymbol{\theta}) &= \eta_1 - \frac{\eta_3}{\beta_3^2} + \eta_{12} \frac{(\mu_j - \bar{\mu})^2}{\beta_2^2 \beta_3^2 \Delta^2} + \eta_{23} \frac{(\mu_j - \bar{\mu})^2}{\beta_2^2 \beta_3^4 \Delta^2}.
\end{aligned}$$

Appendix D: Estimation under multi-lab scenario when all three platforms have data from multiple labs

D.1. Estimation

This section considers the estimates of parameters in the system of ME models under the multi-lab scenario with $k_i > 1$ for $i = 1, 2, 3$. Let \bar{X}_j , \bar{Y}_j and \bar{Z}_j be the average measurements of gene j for the three platforms, and \bar{X}_\cdot , \bar{Y}_\cdot and \bar{Z}_\cdot the overall average measurements of all genes in \mathcal{A} for the three platforms. Let S_{xx} , S_{yy} , S_{zz} , S_{xy} , S_{xz} , S_{yz} be the sample variances and covariances of $\{\bar{X}_j\}_{1 \leq j \leq n}$, $\{\bar{Y}_j\}_{1 \leq j \leq n}$ and $\{\bar{Z}_j\}_{1 \leq j \leq n}$. Following the two-step approach, the estimates of the structural parameters are obtained as follows.

$$\begin{aligned}
\hat{\beta}_2 &= \frac{S_{yz}}{S_{xz}}, \quad \hat{\beta}_3 = \frac{S_{yz}}{S_{xy}}, \quad \hat{\alpha}_2 = \bar{Y}_\cdot - \hat{\beta}_2 \bar{X}_\cdot, \quad \hat{\alpha}_3 = \bar{Z}_\cdot - \hat{\beta}_2 \bar{X}_\cdot, \\
\hat{\psi}_1^2 &= \frac{1}{nk_1} \sum_{j,k} (X_{jk} - \bar{X}_j)^2, \quad \hat{\psi}_2^2 = \frac{1}{nk_2} \sum_{j,k} (Y_{jk} - \bar{Y}_j)^2, \quad \hat{\psi}_3^2 = \frac{1}{nk_3} \sum_{j,k} (Z_{jk} - \bar{Z}_j)^2, \\
\hat{\varsigma}_2^2 &= (S_{xx} - S_{xy}S_{xz}/S_{yz}) - \frac{\hat{\psi}_1^2}{k_1}, \\
\hat{\varsigma}_2^2 &= (S_{yy} - S_{xy}S_{yz}/S_{xz}) - \frac{\hat{\psi}_2^2}{k_2}, \\
\hat{\varsigma}_3^2 &= (S_{zz} - S_{yz}S_{xz}/S_{xy}) - \frac{\hat{\psi}_3^2}{k_3}.
\end{aligned}$$

Let $\boldsymbol{\phi}$ be the vector of the structural parameters, that is, $\boldsymbol{\phi} = (\alpha_2, \alpha_3, \beta_2, \beta_3, \psi^2, \psi_2^2, \psi_3^2, \varsigma_1^2, \varsigma_2^2, \varsigma_3^2)^\top$, and $\hat{\boldsymbol{\phi}}$ the vector of the estimated structural parameters. The following proposition establishes the asymptotic properties of $\hat{\boldsymbol{\phi}}$ under the functional system of ME models with multiple labs as n goes to ∞ .

Proposition 6. Assume the system of ME models (2.10a)-(2.10c) with $k_i \geq 2$ holds for $i = 1, 2, 3$, and the limits in Proposition 1 exist. Then, as n goes to ∞ ,

$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(\mathbf{0}, \Lambda_\phi)$, where \xrightarrow{D} means converge in distribution, and Λ_ϕ is the variance-covariance matrix with its explicit expression given in Section D.2.

Once the estimates of the structural parameters are obtained, we can plug them back into the system of ME models so that the incidental parameters (i.e. μ_j 's) can be estimated by the generalized least squares method. Similar as in Section 2, the best estimate of the incidental parameter μ_j depends on whether j is in \mathcal{A} , $\mathcal{B} - \mathcal{A}$ or $\mathcal{C} - \mathcal{B}$. Define $\lambda_i = \varsigma_i^2 + \psi_i^2/k_i$ for $i = 1, 2, 3$. The estimates of the incidental parameters are given as follows.

$$\hat{\mu}_j^{xyz} = \frac{\bar{X}_{j.}/\hat{\lambda}_1 + \hat{\beta}_2(\bar{Y}_{j.} - \hat{\alpha}_2)/\hat{\lambda}_2 + \hat{\beta}_3(\bar{Z}_{j.} - \hat{\alpha}_3)/\hat{\lambda}_3}{1/\hat{\lambda}_1 + \hat{\beta}_2^2/\hat{\lambda}_2 + \hat{\beta}_3^2/\hat{\lambda}_3}, \quad \text{for } j \in \mathcal{A}, \quad (\text{D.1a})$$

$$\hat{\mu}_j^{yz} = \frac{\hat{\beta}_2(\bar{Y}_{j.} - \hat{\alpha}_2)/\hat{\lambda}_2 + \hat{\beta}_3(\bar{Z}_{j.} - \hat{\alpha}_3)/\hat{\lambda}_3}{\hat{\beta}_2^2/\hat{\lambda}_2 + \hat{\beta}_3^2/\hat{\lambda}_3}, \quad \text{for } j \in \mathcal{B} - \mathcal{A}, \quad (\text{D.1b})$$

$$\hat{\mu}_j^z = \frac{\bar{Z}_{j.} - \hat{\alpha}_3}{\hat{\beta}_3}, \quad \text{for } j \in \mathcal{C} - \mathcal{B}. \quad (\text{D.1c})$$

The variances of the three types of estimated incidental parameters given above have the same forms as their counterparts under the single-lab scenario. The only difference is that σ_i^2 under the single-lab scenario should be replaced by λ_i under the multi-lab scenario. Based on their asymptotic variances, the standard errors of the estimates can be similarly obtained, which are denoted as $\sqrt{\hat{\gamma}_{xyz}}$, $\sqrt{\hat{\gamma}_{yz}}$ and $\sqrt{\hat{\gamma}_z}$. The estimates $\{\hat{\mu}_j^{xyz} : j \in \mathcal{A}\}$, $\{\hat{\mu}_j^{yz} : j \in \mathcal{B} - \mathcal{A}\}$ and $\{\hat{\mu}_j^z : j \in \mathcal{C} - \mathcal{B}\}$, together with their standard errors, can be used in downstream gene expression analysis.

D.2. Asymptotic Variances of the Estimates of Structural Parameters under multi-lab scenario when all three platforms have data from multiple labs

Recall that $\Delta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\mu_j - \bar{\mu})^2$. Denote $\lambda_i = \varsigma_i^2 + \psi_i^2/k_i$ for $i = 1, 2, 3$. Denote $\lambda_{ij} = \lambda_i \lambda_j$ for $i, j = 1, 2, 3$. The explicit form of Λ_ϕ is given as follows.

$$\Lambda_\phi = \begin{bmatrix} \Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} & \Lambda_{\alpha\varsigma} & \mathbf{0} \\ \Lambda_{\alpha\beta}^\top & \Lambda_{\beta\beta} & \Lambda_{\beta\varsigma} & \mathbf{0} \\ \Lambda_{\alpha\varsigma}^\top & \Lambda_{\beta\varsigma}^\top & \Lambda_{\varsigma\varsigma} & \Lambda_{\varsigma\psi} \\ \mathbf{0} & \mathbf{0} & \Lambda_{\varsigma\psi}^\top & \Lambda_{\psi\psi} \end{bmatrix}$$

where

$$\begin{aligned}
\Lambda_{\beta\beta} &= \frac{1}{\Delta^2} \begin{bmatrix} \frac{\beta_2^2}{\beta_3^2} \lambda_{13} + \frac{1}{\beta_3^2} \lambda_{23} & \frac{\lambda_{23}}{\beta_2 \beta_3} \\ \frac{\lambda_{23}}{\beta_2 \beta_3} & \frac{\beta_3^2}{\beta_2^2} \lambda_{12} + \frac{1}{\beta_2^2} \lambda_{23} \end{bmatrix}, \\
\Lambda_{\alpha\alpha} &= \bar{\mu}^2 \Lambda_{\beta\beta} + \begin{bmatrix} \beta_2^2 \lambda_1 + \lambda_2 & \beta_2 \beta_3 \lambda_1 \\ \beta_2 \beta_3 \lambda_1 & \beta_3^2 \lambda_1 + \lambda_3 \end{bmatrix}, \\
\Lambda_{\alpha\beta} &= -\bar{\mu} \Lambda_{\beta\beta}, \\
\Lambda_{\beta\varsigma} &= \frac{1}{\Delta} \begin{bmatrix} \frac{\beta_2}{\beta_3^2} \lambda_{13} + \frac{1}{\beta_2 \beta_3^2} \lambda_{23} & -\left(\frac{\beta_3^2}{\beta_2^2} \lambda_{13} + \frac{\beta_2}{\beta_3^2} \lambda_{23}\right) & \beta_2 \lambda_{13} - \frac{1}{\beta_2} \lambda_{23} \\ \frac{\beta_3}{\beta_2^2} \lambda_{12} + \frac{1}{\beta_2^2 \beta_3} \lambda_{23} & \beta_3 \lambda_{12} - \frac{1}{\beta_3} \lambda_{23} & -\left(\frac{\beta_3^2}{\beta_2^2} \lambda_{12} + \frac{\beta_3}{\beta_2^2} \lambda_{23}\right) \end{bmatrix}, \\
\Lambda_{\alpha\varsigma} &= -\bar{\mu} \Lambda_{\beta\varsigma}, \\
\Lambda_{\varsigma\varsigma} &= \text{diag} \left(2\lambda_1^2 + \frac{2\psi_1^4}{k_1^3}, 2\lambda_2^2 + \frac{2\psi_2^4}{k_2^3}, 2\lambda_3^2 + \frac{2\psi_3^4}{k_3^3} \right) \\
&\quad + \begin{bmatrix} \frac{\lambda_{12}}{\beta_2^2} + \frac{\lambda_{13}}{\beta_3^2} + \frac{\lambda_{23}}{\beta_2^2 \beta_3^2} & \lambda_{12} - \frac{\beta_2^2}{\beta_3^2} \lambda_{13} - \frac{\lambda_{23}}{\beta_2^2} & \lambda_{13} - \frac{\beta_3^2}{\beta_2^2} \lambda_{12} - \frac{\lambda_{23}}{\beta_2^2} \\ \lambda_{12} - \frac{\beta_2^2}{\beta_3^2} \lambda_{13} - \frac{\lambda_{23}}{\beta_2^2} & \beta_2^2 \lambda_{12} + \frac{\beta_3^2}{\beta_2^2} \lambda_{13} + \frac{\beta_2^2}{\beta_3^2} \lambda_{23} & \lambda_{23} - \beta_3^2 \lambda_{12} - \beta_2^2 \lambda_{13} \\ \lambda_{13} - \frac{\beta_3^2}{\beta_2^2} \lambda_{12} - \frac{\lambda_{23}}{\beta_2^2} & \lambda_{23} - \beta_3^2 \lambda_{12} - \beta_2^2 \lambda_{13} & \beta_3^2 \lambda_{13} + \frac{\beta_4^2}{\beta_2^2} \lambda_{12} + \frac{\beta_3^2}{\beta_2^2} \lambda_{23} \end{bmatrix}, \\
\Lambda_{\varsigma\psi} &= \text{diag} \left(-\frac{2\psi_1^4}{k_1^2}, -\frac{2\psi_2^4}{k_2^2}, -\frac{2\psi_3^4}{k_3^2} \right), \\
\Lambda_{\psi\psi} &= \text{diag} \left(\frac{2\psi_1^4}{k_1}, \frac{2\psi_2^4}{k_2}, \frac{2\psi_3^4}{k_3} \right).
\end{aligned}$$

Appendix E: Asymptotic variances of the estimates of structural parameters under Multi-lab scenario when the qRT-PCR platform has data from one lab

Recall that $\Delta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\mu_j - \bar{\mu})^2$. Denote $\lambda_1 = \sigma_1^2$, $\lambda_i = \varsigma_i^2 + \psi_i^2/k_i$ for $i = 2, 3$. Denote $\lambda_{ij} = \lambda_i \lambda_j$ for $i, j = 1, 2, 3$. The explicit form of Λ_ϕ is given as follows.

$$\Lambda_\phi = \begin{bmatrix} \Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} & \Lambda_{\alpha\varsigma} & \mathbf{0} \\ \Lambda_{\alpha\beta}^\top & \Lambda_{\beta\beta} & \Lambda_{\beta\varsigma} & \mathbf{0} \\ \Lambda_{\alpha\varsigma}^\top & \Lambda_{\beta\varsigma}^\top & \Lambda_{\varsigma\varsigma} & \Lambda_{\varsigma\psi} \\ \mathbf{0} & \mathbf{0} & \Lambda_{\varsigma\psi}^\top & \Lambda_{\psi\psi} \end{bmatrix}$$

where

$$\begin{aligned}
\Lambda_{\beta\beta} &= \frac{1}{\Delta^2} \begin{bmatrix} \frac{\beta_2^2}{\beta_3^2}\lambda_{13} + \frac{1}{\beta_3^2}\lambda_{23} & \frac{\lambda_{23}}{\beta_2\beta_3} \\ \frac{\lambda_{23}}{\beta_2\beta_3} & \frac{\beta_3^2}{\beta_2^2}\lambda_{12} + \frac{1}{\beta_2^2}\lambda_{23} \end{bmatrix}, \\
\Lambda_{\alpha\alpha} &= \bar{\mu}^2 \Lambda_{\beta\beta} + \begin{bmatrix} \beta_2^2\lambda_1 + \lambda_2 & \beta_2\beta_3\lambda_1 \\ \beta_2\beta_3\lambda_1 & \beta_3^2\lambda_1 + \lambda_3 \end{bmatrix}, \\
\Lambda_{\alpha\beta} &= -\bar{\mu}\Lambda_{\beta\beta}, \\
\Lambda_{\beta\varsigma} &= \frac{1}{\Delta} \begin{bmatrix} \frac{\beta_2}{\beta_3^2}\lambda_{13} + \frac{1}{\beta_2\beta_3^2}\lambda_{23} & -\left(\frac{\beta_3^3}{\beta_2^2}\lambda_{13} + \frac{\beta_2}{\beta_3^2}\lambda_{23}\right) & \beta_2\lambda_{13} - \frac{1}{\beta_2}\lambda_{23} \\ \frac{\beta_3}{\beta_2^2}\lambda_{12} + \frac{1}{\beta_2^2\beta_3}\lambda_{23} & \beta_3\lambda_{12} - \frac{1}{\beta_3}\lambda_{23} & -\left(\frac{\beta_3^2}{\beta_2^2}\lambda_{12} + \frac{\beta_3}{\beta_2^2}\lambda_{23}\right) \end{bmatrix}, \\
\Lambda_{\alpha\varsigma} &= -\bar{\mu}\Lambda_{\beta\varsigma}, \\
\Lambda_{\varsigma\varsigma} &= \text{diag}\left(2\lambda_1^2, 2\lambda_2^2 + \frac{2\psi_2^4}{k_2^3}, 2\lambda_3^2 + \frac{2\psi_3^4}{k_3^3}\right) \\
&\quad + \begin{bmatrix} \frac{\lambda_{12}}{\beta_2^2} + \frac{\lambda_{13}}{\beta_3^2} + \frac{\lambda_{23}}{\beta_2^2\beta_3^2} & \lambda_{12} - \frac{\beta_2^2}{\beta_3^2}\lambda_{13} - \frac{\lambda_{23}}{\beta_2^2} & \lambda_{13} - \frac{\beta_3^2}{\beta_2^2}\lambda_{12} - \frac{\lambda_{23}}{\beta_2^2} \\ \lambda_{12} - \frac{\beta_2^2}{\beta_3^2}\lambda_{13} - \frac{\lambda_{23}}{\beta_2^2} & \beta_2^2\lambda_{12} + \frac{\beta_2}{\beta_3^2}\lambda_{13} + \frac{\beta_2^2}{\beta_3^2}\lambda_{23} & \lambda_{23} - \beta_3^2\lambda_{12} - \beta_2^2\lambda_{13} \\ \lambda_{13} - \frac{\beta_3^2}{\beta_2^2}\lambda_{12} - \frac{\lambda_{23}}{\beta_2^2} & \lambda_{23} - \beta_3^2\lambda_{12} - \beta_2^2\lambda_{13} & \beta_3^2\lambda_{13} + \frac{\beta_3^4}{\beta_2^2}\lambda_{12} + \frac{\beta_3^2}{\beta_2^2}\lambda_{23} \end{bmatrix}, \\
\Lambda_{\varsigma\psi} &= \text{diag}\left(-\frac{2\psi_2^4}{k_2^2}, -\frac{2\psi_3^4}{k_3^2}\right), \\
\Lambda_{\psi\psi} &= \text{diag}\left(\frac{2\psi_2^4}{k_2}, \frac{2\psi_3^4}{k_3}\right).
\end{aligned}$$

Appendix F: Simulation Results of Structural Parameters under Single-Lab Scenario

As discussed in Section 2, the structural parameters are fundamentally different from the incidental parameters, in that the former can be consistently estimated as the sample size n goes to infinity. In other words, as n increases, the estimates of the structural parameters will converge to their respective targets as shown in Proposition 1. Figure 5 shows the MSEs of $\hat{\beta}_2$ under the three model settings in Plot (a) and the MSEs of $\hat{\beta}_3$ under the three model settings in Plot (b) as n increases. Figure 6 demonstrates the MSEs of $\hat{\sigma}_i^2$ ($i = 1, 2, 3$) under the three model settings in Plots (a), (b), and (c), respectively. It is clear that, as n increases, all the MSEs decrease towards zero, indicating the convergence of the corresponding estimates.

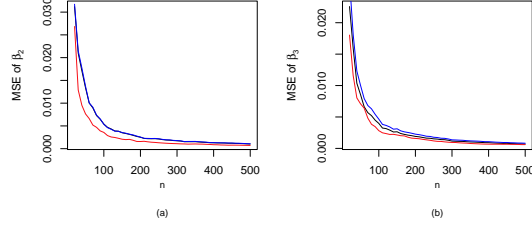
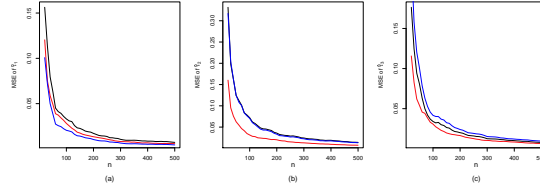


Fig 5: MSE of Estimated Slopes

Plot (a) shows the MSE of $\hat{\beta}_2$ and Plot (b) shows the MSE of $\hat{\beta}_3$ as n increases. The solid, dashed and dotted lines correspond to model setting 1, 2, and 3 in simulation study.

Fig 6: MSE of $\hat{\sigma}_i^2$

Plot (a) shows the MSE of $\hat{\sigma}_1^2$, Plot (b) shows the MSE of $\hat{\sigma}_2^2$ and Plot (c) shows the MSE of $\hat{\sigma}_3^2$ as n increases. The solid, dashed and dotted lines correspond to model setting 1, 2, and 3 in simulation study.

Appendix G: RNA-Seq Experiments and Data Preprocessing

The procedure of a RNA-Seq experiment is described as follows. First, the total RNA sample is purified to mRNA and further fragmented into small segments. These mRNA fragments are converted to cDNA by reverse transcription, and go through end repair, adapter ligation, size selection and amplification steps, resulting in a library of cDNA fragments. The cDNA fragments are then sequenced by a sequencing machine, generating millions of reads (i.e. nucleotide sequence) of a certain length. Short reads are then mapped back to a reference genome, and are summarized as a sequence of counts. The counts at a nucleotide position is the number of reads that start at this position. Various normalization methods can be used to quantify gene expression level based on the counts data. In a RNA-Seq experiment using Illumina Genome Analyzer, multiple lanes (up to 7) can be used to sequence cDNA fragments and generate short reads, and the used lanes are treated as technical replicates. In this article, we used the average value of RPKM measurements (in log-2 scale) across the technical replicates (lanes) as the expression measurement for each gene.

For the RNA-Seq data used in this paper, SRA010153 contains short reads of 35 base pairs from 7 lanes for both UHR and Brain samples, whereas SRA008403 contains short reads of 31 base pairs from 7 lanes for the Brain sample but only from 3 lanes for the UHR sample. To avoid ambiguity caused by the unbalance in the number of lanes for the UHR sample, we use the first 3 lanes in SRA010153 for the UHR sample. The short reads in both SRA010153 and SRA008403 were aligned to the UCSC hg19 reference genome using Bowtie 0.12.7 (?). Allowing up to two mismatches per read and counting only uniquely mapped reads, there resulted in about 2-4 million mapped reads per lane for SRA010153 and about 0.7-0.8 million reads per lane for SRA008403. Using the Reference Sequence (RefSeq) database, the mapped reads were used to calculate the RPKM value for each gene.

Appendix H: Residual Plots and Plots of Box-Cox Transformation

The system of ME models (2.1a)-(2.1c) imposes the homoscedasticity and normality assumption on the measurement errors for each platform. In order to check the homoscedasticity assumption, we generated residual plots and constructed approximate 95% confidence intervals for Box-Cox transformations. The three plots in Figure 7 from left to right are the plots of residuals versus $\hat{\mu}_j$ for qRT-PCR, microarray and RNA-Seq in the UHR sample, respectively. The three plots do not demonstrate strong heteroscedastic patterns. The three plots in Figure 8 from left to right present the 95% confidence interval for the Box-Cox transformation based on the residuals for qRT-PCR, microarray and RNA-Seq in the UHR sample. The 95% confidence intervals for Box-Cox transformations corresponding to the qRT-PCR and RNA-Seq platforms both contain 1, and 1 is on the boundary of the 95% confidence interval corresponding to the microarray platform. These results indicate that there does not exist significant violation of the homoscedasticity assumption on the platform measurement errors.

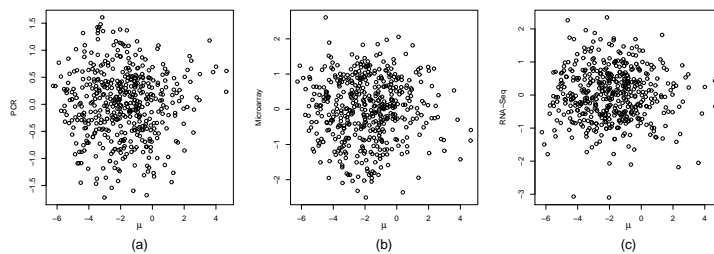


Fig 7: Plots of platform residuals for qRT-PCR (a), microarray (b) and RNA-Seq (c) based on the measurements in the UHR sample

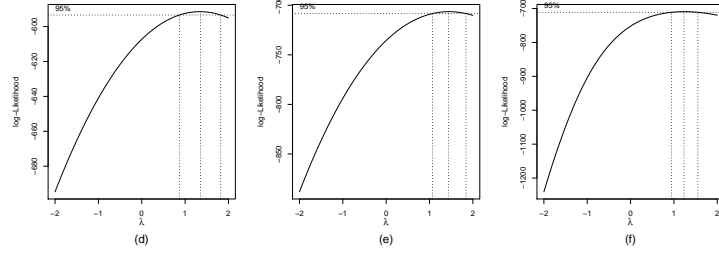


Fig 8: Plots from Box-Cox test On the Residuals for qRT-PCR (d), microarray (e) and RNA-Seq (f) based on the measurements in the UHR sample

Appendix I: Results under Multi-Lab Scenario

In this section, we applied the system of ME models for the multi-lab scenario to analyze the entire expression data of the UHR and Brain samples. The benefit of using the entire data instead of a combination of the three types of data from single labs is two-fold. First, due to the inclusion of RNA-Seq and microarray data from multiple labs, we were able to decompose the variance component due to a platform into the variance component due to the technology of the platform and the variance component due to the involved labs, respectively, that is, $\sigma_i^2 = \psi_i^2 + \zeta_i^2$ for $i = 2, 3$. Second, the estimates of the structural parameters and the calibrated measurements become more accurate with reduced standard errors.

The estimates of structural parameters, together with their standard errors, are given in Table 4. We examine the estimated variance components of the RNA-Seq platform first. For the RNA-Seq platform, $\hat{\zeta}_3^2 = 0.8063$ and $\hat{\psi}_3^2 = 0.1902$ in the Brain sample; $\hat{\zeta}_3^2 = 0.7422$ and $\hat{\psi}_3^2 = 0.0687$ in the UHR sample. The variance components due to the RNA-Seq technology are about 5 and 10 folds of the variance components due to the labs in the Brain and UHR samples, respectively. Therefore, the measurement errors in RNA-Seq measurements are mainly dominated by the measurement error due to the RNA-Seq technology. The sum of $\hat{\zeta}_3^2$ and $\hat{\psi}_3^2$ can be used as an estimate of σ_3^2 , which is the variance of the RNA-Seq platform, and we denote the resulting estimate as $\tilde{\sigma}_3^2$. In the Brain sample, $\tilde{\sigma}_3^2 = 0.9965$, and in the UHR sample, $\tilde{\sigma}_3^2 = 0.8109$, which are not significantly different from the values of $\hat{\sigma}_3^2$ for the two RNA samples reported in Table 2 (Section 4.2.2 in the main article).

For the microarray platform, $\hat{\zeta}_2^2 = 1.3190$ and $\hat{\psi}_2^2 = 0.0185$ in the Brain sample, and $\hat{\zeta}_2^2 = 1.0388$ and $\hat{\psi}_2^2 = 0.0193$ in the UHR sample. Similar to the RNA-Seq measurements, the variance component due the microarray technology dominates the variance components due to the involved labs. The sum of the two variance components estimated under the multi-lab scenario is not significantly different from the variance component due to the combination of the microarray technology and lab estimated under the single-lab scenario, that is, $\hat{\psi}_2^2 + \hat{\zeta}_2^2 = \hat{\sigma}_2^2$,

TABLE 4
Estimates of Structural Parameters Using the Entire UHR and Brain Dataset

	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$		$\hat{\sigma}_3^2$	
						$\hat{\zeta}_2^2$	$\hat{\psi}_2^2$	$\hat{\zeta}_3^2$	$\hat{\psi}_3^2$
Brain	8.7822 (0.0708)	4.8970 (0.0809)	0.7416 (0.0198)	1.0001 (0.0314)	0.7135 (0.1117)	1.3190 (0.1018)	0.0185 (0.0039)	0.8063 (0.1175)	0.1902 (0.0092)
UHR	9.1014 (0.0589)	5.4507 (0.0661)	0.7689 (0.0131)	1.0151 (0.0203)	0.6940 (0.0911)	1.0388 (0.0796)	0.0193 (0.0040)	0.7422 (0.0943)	0.0687 (0.0032)

in both the Brain and UHR samples.

After the estimates of the structural parameters were obtained, we used the three types of calibrated estimates of the incidental parameters to calibrate the gene expression levels and calculated their standard errors. The calibrated expression values based on the entire data are of a higher quality than those calculated in the single-lab scenario. These calibrated expression level measurements, together with their standard errors, can be obtained from the authors upon request.